

•

**Computational, biochemical
and NMR-driven
structural studies on
nucleosomal DNA**



•

Clara van Emmerik

Computational, biochemical and NMR-driven structural studies on nucleosomal DNA

Clara L. van Emmerik

Doctoral thesis

Computational, biochemical and NMR-driven structural studies
on nucleosomal DNA

C.L. van Emmerik

ISBN: 978-94-6380-673-2

Cover design by Franke Margrete | www.frankemargrete.nl

Printed by ProefschriftMaken | www.proefschriftmaken.nl

Copyright © 2020 Clara van Emmerik

Computational, biochemical and NMR-driven structural studies on nucleosomal DNA

Computationale, biochemische en NMR-gedreven structuurstudies aan nucleosomaal DNA

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 10 februari 2020 des middags te 2.30 uur

door

Clara Louise van Emmerik

geboren op 5 november 1989
te Zwolle

Promotor:

Prof.dr. A.M.J.J. Bonvin

Copromotor:

Dr. H. Van Ingen

TABLE OF CONTENTS

List of Abbreviations	6
1. Introduction	9
2. Architectural principles of archaeal and eukaryotic chromatin complexes.....	65
3. Ramified rolling circle amplification for synthesis of nucleosomal DNA sequences.....	95
4. The role of the N-terminal region in autoregulation of nucleosome remodeling by ISWI .	119
5. Segmental one-strand isotope-labeling of nucleosomal DNA-601 as a tool in NMR studies of nucleosome interactions.....	151
6. General Discussion	175
English Summary	187
Nederlandse Samenvatting	190
Acknowledgements	195
Curriculum Vitae.....	199
List of Publications	199

LIST OF ABBREVIATIONS

4HB: four helix bundle

ADP: adenosine diphosphate

ATP: adenosine triphosphate

BME: β -mercaptoethanol

bp: base pair

CD: circular dichroism

CEST: chemical exchange saturation transfer

CP: cross-polarization

CPMG: Carr-Purcell-Meiboom-Gill sequence

Cryo-EM: cryo-electron microscopy

CSP: chemical shift perturbation

DNA: deoxyribonucleic acid

dNTPs: deoxyribose nucleoside triphosphates

dsDNA: double-stranded DNA

DTT: dithiothreitol

EDTA: ethylenediaminetetraacetic acid

EM: electron microscopy

FA: fluorescence anisotropy

HMG: high mobility group proteins

HMQC: heteronuclear multiple quantum correlation

HSQC: heteronuclear single quantum correlation

INEPT: insensitive nuclei enhanced by polarization transfer

IPTG: isopropyl β -D-1-thiogalactopyranoside

MAS: magic angle spinning

MeTROSY: methyl transverse relaxation optimized spectroscopy

MNase: micrococcal nuclease

MWCO: molecular weight cut-off

NCP: nucleosome core particle

NMR: nuclear magnetic resonance

NOESY: nuclear Overhauser effect spectroscopy

nt: nucleotide

PAGE: polyacrylamide gel electrophoresis

PCR: polymerase chain reaction
PCS: pseudo-contact shift
PDB: Protein Data Bank
PMSF: phenylmethylsulfonyl fluoride
PRE: paramagnetic relaxation enhancement
PTM: post-translational modification
RDC: residual dipolar coupling
RNA: ribonucleic acid
rRCA: ramified rolling circle amplification
SAXS: small angle X-ray scattering
ssDNA: single-stranded DNA
ssNMR: solid-state nuclear magnetic resonance
TPM: tethered particle motion
TOCSY: total correlation spectroscopy

1

INTRODUCTION

Based on the review article: **C.L. van Emmerik** and H. van Ingen (2019) Unspinning chromatin: revealing the dynamic nucleosome landscape by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **110**, 1–19

ABSTRACT

NMR is an essential technique for obtaining information at atomic resolution on the structure, motions and interactions of biomolecules. Here, we review the contribution of NMR to our understanding of the fundamental unit of chromatin: the nucleosome. Nucleosomes compact the genome by wrapping the DNA around a protein core, the histone octamer, thereby protecting genomic integrity. Crucially, the imposed barrier also allows strict regulation of gene expression, DNA replication and DNA repair processes through an intricate system of histone and DNA modifications and a wide range of interactions between nucleosomes and chromatin factors. In this review, we describe how NMR has contributed to deciphering the molecular basis of nucleosome function. Starting from pioneering studies in the 1960's using natural abundance NMR studies, we focus on the progress in sample preparation and NMR methodology that has allowed high-resolution studies on the nucleosome and its subunits. We summarize the results and approaches of state-of-the-art NMR studies on nucleosomal DNA, histone complexes, nucleosomes and nucleosomal arrays. These studies highlight the particular strength of NMR in studying nucleosome dynamics and nucleosome-protein interactions. Finally, we look ahead to exciting new possibilities that will be afforded by on-going developments in solution and solid-state NMR. By increasing both the depth and breadth of nucleosome NMR studies, it will be possible to offer a unique perspective on the dynamic landscape of nucleosomes and its interacting proteins.

1. INTRODUCTION

To understand the molecular basis of life requires intimate knowledge of the structures and motions of biomolecules, as well as the interactions between them. NMR has proven to be uniquely and extremely sensitive to these three fundamental aspects of biomolecules and has thus become a powerful tool in structural biology. NMR has made key contributions in diverse areas of structural biology ranging from protein-protein interactions, protein-RNA complexes and protein folding to membrane proteins (1-8). Focusing on the cell nucleus, NMR has been instrumental in advancing our understanding of protein-DNA interactions that control gene transcription (9). Here, we focus on the packaging of DNA into chromatin, which represents one of the most fundamental layers of cell biology. Chromatin provides the required structural compaction of DNA to fit in the nucleus and plays crucial roles in controlling cell fate and protecting genome integrity. Chromatin function ultimately depends on the nucleosome, its repeating unit, and thus on the structural and dynamical properties of the nucleosome and its interactions with a wide range of chromatin factors.

The high-resolution structure of the nucleosome was solved by crystallography over 20 years ago (10), offering the first detailed insights into its molecular organization. Ever since, many structures of nucleosomes have been solved, mostly using different histone species, mutants or complexes with small molecules. Essentially the same nucleosome structure has been found in every case. Yet, it is clear that there is a wide and dynamic 'nucleosome landscape': nucleosomes vary in DNA sequence, histone protein composition, as well as epigenetic status, as defined by the covalent modifications attached to DNA and histone proteins. To capture this diversity and sample the complete landscape is a major challenge. When including the many possible nucleosome-protein interactions in this conformational landscape, the challenge is only exacerbated, and this is reflected in the very few structures of nucleosome-protein complexes that have been solved. Moreover, ever since the demonstration that nucleosomes are inherently dynamic assemblies (11) and contain highly disordered histone tails (12), it remains a key question how nucleosome dynamics and the presence of such flexible parts contribute to nucleosome assembly or disassembly, the binding of chromatin factors, and the remodeling of chromatin structure. Finally, nucleosomes are organized in a higher-order chromatin structure, in

which they interact with each other and linker histones in a yet unknown and dynamic manner, further expanding the conformational landscape of nucleosomes.

Here, we review the ways in which NMR has met and overcome the challenges in characterizing the structure, dynamics and interactions of nucleosomes. After a brief overview of nucleosome structure and function, we survey the pioneering NMR studies starting in the late sixties of the last century. We then describe the state-of-the-art solution and solid-state NMR (ssNMR) methods that have enabled the high-resolution study of the nucleosome by NMR, and outline the required isotope labeling and sample preparation. The two penultimate sections focus on the application of NMR in studies of the nucleosome and nucleosome-protein interactions. Concluding our review, we will look forward to the exciting opportunities that lie ahead to capture the dynamic nucleosome landscape by NMR.

2. OVERVIEW OF NUCLEOSOME STRUCTURE AND FUNCTION

Nucleosomes act as the genome's guardian, protecting its integrity by wrapping the DNA around histone octamers and compacting it into higher order chromatin. At the same time, they act as gate keepers by regulating the binding of proteins that carry out DNA-templated processes like transcription, replication and repair. Tight regulation of these fundamental processes is imposed through the structural conformation and modification state of nucleosomes, and is essential for a healthy cell.

The nucleosome core particle (NCP) consists of two copies of histones H2A, H2B, H3 and H4 and accommodates ~147 bp of DNA, wrapped in 1.6 left-handed superhelical turns around the histone octamer (Fig. 1, top left) (10). This 200 kDa assembly has an approximate disk-like shape with a pseudo-twofold symmetry axis, the dyad, going through the central base pair of the DNA. NCPs that include a varying length of linker DNA are called nucleosomes, whereas binding of the linker histone H1 to the linker DNA forms a chromatosome (13).

The core histones each consist of a long α helix, enclosed by two shorter ones, and a flexible, lysine-rich N-terminal tail (Fig. 1). The histones fold into H2A-H2B and H3-H4 heterodimers forming the so-called 'handshake motif' through extensive hydrophobic interactions between their histone fold domains. H3-H4 dimers form a tetramer and associate with two H2A-H2B dimers to complete the histone octamer. The DNA around

it is held in place primarily by histone arginine side chains that penetrate into the DNA minor groove approximately every helical turn, forming three contact points per histone dimer. As extreme bending of the helix is required at these minor grooves, AT base pairs are more easily accommodated there than the more rigid GC base pairs. A sequence with appropriately alternating AT- and GC-rich regions will therefore enforce formation of a nucleosome with a well-defined position. This is the basis of so-called strong-positioning sequences like the Widom 601 sequence (14) that are frequently used for *in vitro* nucleosome reconstitution.

Nucleosomes are the crucial binding platforms for a large variety of proteins and protein complexes that control chromatin biology and define the functional state of chromatin (Fig. 1, bottom). These include architectural proteins, such as the abundant nucleosomal high mobility group proteins (HMGNs) (15), pioneer transcription factors, i.e. those that are capable of binding target DNA sites that are embedded in a nucleosome (16), and an array of histone modifying enzymes that install or remove post-translational modifications (PTMs) (17,18). These modifications, which can loosely be referred to as epigenetic marks, can have a direct effect on nucleosome stability or chromatin structure. Many such marks serve to recruit chromatin factors to a distinct genomic location. This recruitment relies on the specific binding of dedicated 'reader' domains to the modification, or a specific combination of modifications. In addition, an important class of proteins can remodel nucleosome and chromatin structure. Chromatin remodeling enzymes convert the energy of ATP to move nucleosomes along the DNA (19,20), while histone chaperones guide the assembly and disassembly of nucleosomes, and the incorporation of histone variants. These non-canonical histones have slightly different properties from their canonical counterparts, changing the structure or stability of the nucleosomes (21,22).

The tremendous breadth of the nucleosome interaction landscape is highlighted by a recent mass spectrometry-based survey that identified hundreds of protein-nucleosome interactions based on *in situ* crosslinking experiments (23). Nucleosome-binding proteins can bind to the histone tails, the DNA, and also the nucleosome core. Several proteins have been identified that use a single arginine to anchor to the so-called acidic patch, formed by residues from histone H2A and H2B on the core surface of the octamer (24,25) (Fig. 1). Importantly, it is more and more appreciated that many proteins interacting with the nucleosome bind to multiple regions simultaneously,

including the nucleosomal DNA, increasing their binding affinity and specificity in the nucleosomal context.

On a larger scale, nucleosomes interacting with each other facilitate the assembly into higher order chromatin structures (Fig. 1, top right). These structures can include chromatin loops, such as CTCF stabilized loops from an enhancer to a promoter (26), and highly condensed fiber structures (27). Chromosomes in addition have an internal structure, referred to as topologically associated domains (TADs) consisting of regions within the chromosome that are often in close contact (28). In addition, phase-separation mechanisms could play a role in chromatin organization, exemplified by the formation of heterochromatin droplets in the presence of phosphorylated heterochromatin protein 1 α (HP1 α) (29). The largest structural assemblies within the nucleus are chromosome territories, which have changeable boundaries and are separated by interchromatin compartments (30).

3. PIONEERING NMR STUDIES ON THE NUCLEOSOME

Soon after the discovery of histones and nucleosomes as building blocks of chromatin (see Table 1 for an overview of the historic milestones in structural studies of chromatin), NMR was applied to probe their conformational landscape. The earliest studies used histones purified from chicken erythrocytes or calf thymus chromatin by acid and ethanol fractionation (31-33). Later, it became possible to isolate native histone dimer and tetramer complexes from 'chromatin gel' by protamine replacement of the histones in combination with size exclusion chromatography under non-denaturing conditions (34,35). Nucleosome core particles could be obtained by micrococcal nuclease (MNase) digestion of isolated chromatin (36). The samples obtained were obviously inhomogeneous, due to varying DNA sequences, presence of histone variants and PTMs, and possibly associated proteins such as linker histones. As isotope labeling of these materials was not possible, studies were limited to natural abundance NMR experiments, observing ^1H , ^{31}P , or ^{13}C nuclei.

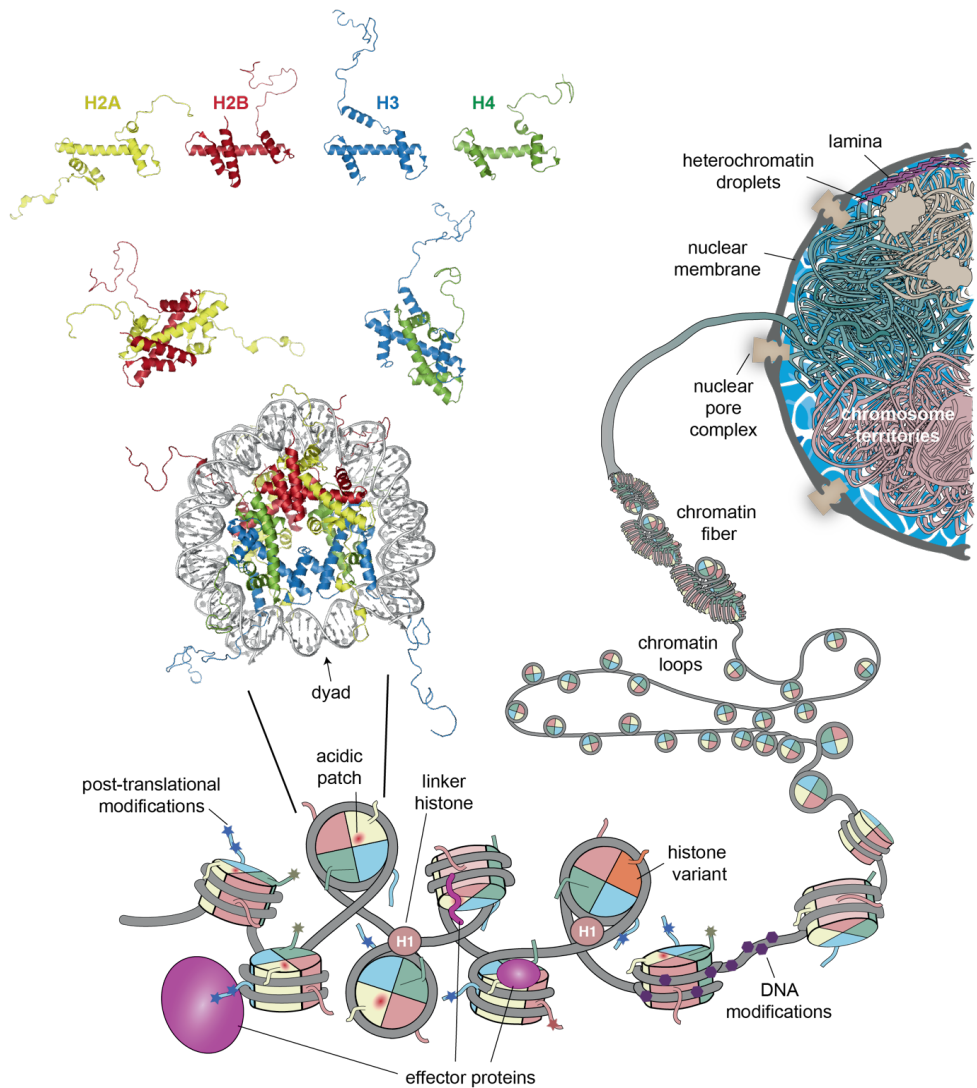


Figure 1. Hierarchical view of nucleosome and chromatin organization. Histone, histone dimer and nucleosome structures are depicted in cartoon representation in the top left corner, showing histone H3 in blue, H4 in green, H2A in yellow, H2B in red and the DNA in grey. The dyad base pair is indicated by the arrow. The bottom of the figure shows schematically the different modifications and interactions of nucleosomes. The right side of the figure illustrates the compaction into higher order chromatin structures within the cell nucleus.

3.1. REVEALING THE NATURE OF HISTONES

Histone proteins were discovered in goose blood cells by Albrecht Kossel in 1884 (37). They were characterized as a very diverse group of proteins rich in basic residues and likely involved in gene regulation. In 1965, fractionation revealed five histone fractions termed F1, F2A1, F2A2, F2B and F3 (38), which were later renamed histones H1, H4, H2A, H2B and H3, respectively. Histones were isolated under denaturing conditions and refolded by dialysis to water and increasing the salt molarity, which also promotes histone aggregation. Early NMR studies used 1D ^1H spectra to deduce the helical segments of individual histones and study their folding and aggregation. The first example is a study by the late Morton Bradbury, who pioneered the application of NMR to study chromatin and whose extensive body of work anticipated much of the research that is still on-going today. This work from 1967 showed that alanine, leucine, isoleucine and valine are part of the folded, helical segment (39). As the primary sequences of the histones became available (40-43), several studies used this information combined with broadening of aromatic and apolar resonances upon histone folding to conclude that the C-terminal part of the histones is helical and that the N-terminal lysine-rich region remains mobile (Fig. 2A) (44-51). It was shown that the apolar side of the helical segment facilitates interactions between histones of the same type, while the random coil segment remains mobile in the aggregate (44,45). Using highly concentrated samples with protein concentrations up to 100 mg/mL, the large chemical shift dispersion of ^{13}C NMR was exploited to show that aggregation does not affect the secondary structure (52-54). Of note, one report employed a methylation/demethylation cycle to obtain ^{13}C -labeled methionine methyl groups to study the self-aggregation of H2B, a first attempt in isotope-labeling (55).

After the milestone discoveries by Thomas and Kornberg that histones associate pairwise, H3 to H4 and H2A to H2B, and that these pairwise interactions occur in the repeating unit of chromatin (56,57), the attention shifted from self-aggregation of histones to their native interactions and tertiary structure. In 1976, Moss *et al.* (58,59) isolated native H3-H4 and H2A-H2B complexes from calf thymus and showed by 1D ^1H NMR that these have a specific tertiary fold as evidenced by ring-current upfield shifted methyl peaks, and that, while folded, they still contain a large, unstructured region, likely corresponding to the N-terminal region (Fig. 2B). The folding of H3 in refolded histone octamer mixtures was studied by ^{19}F NMR using trifluoro-acetylation

of the native cysteines in calf H3 confirming the presence of defined tertiary structures at intermediate ionic strength (60). To map the interaction surface in native H3-H4 tetramers, mixtures of H3 and H4 fragments were studied by ^1H NMR showing that the first ~40 residues of both histones are not required for complex formation (61). Most lysines and alanines, located in the N-terminal region, were shown to remain in random coil state in several different histone complexes (12,62,63). In the mid 1970's, work from Lilley *et al.* showed by NMR relaxation measurements that these regions are indeed highly flexible and pushed the notion of 'histone tails' for the first time (12,63).

Next, the role of these tails in the native nucleosome was addressed. Early reports had proposed that the very basic tails would provide the fundamental contacts between histones and the negatively charged DNA (64). However, a ^1H NMR study on the salt-induced dissociation of nucleosomes and tetrasomes argued that only the H3 and H4 tails and not the H2A-H2B tails are bound to the DNA (65). In addition, trypsinized nucleosomes, lacking the basic tails, were shown to retain a DNA-bound histone core (66). The Bradbury group showed that H4 tail peptides only weakly bind DNA, and this interaction is abolished upon acetylation, a first example of the study of histone PTMs (67). Later studies showed that the histone tails, although not essential for histone-DNA binding and highly mobile even in oligonucleosomes, do interact with the DNA under physiological salt conditions, but much more weakly than do the histone cores (68-70). The basic tails are suggested to shield neighboring DNA charges in this way, thereby stabilizing higher order chromatin packing.

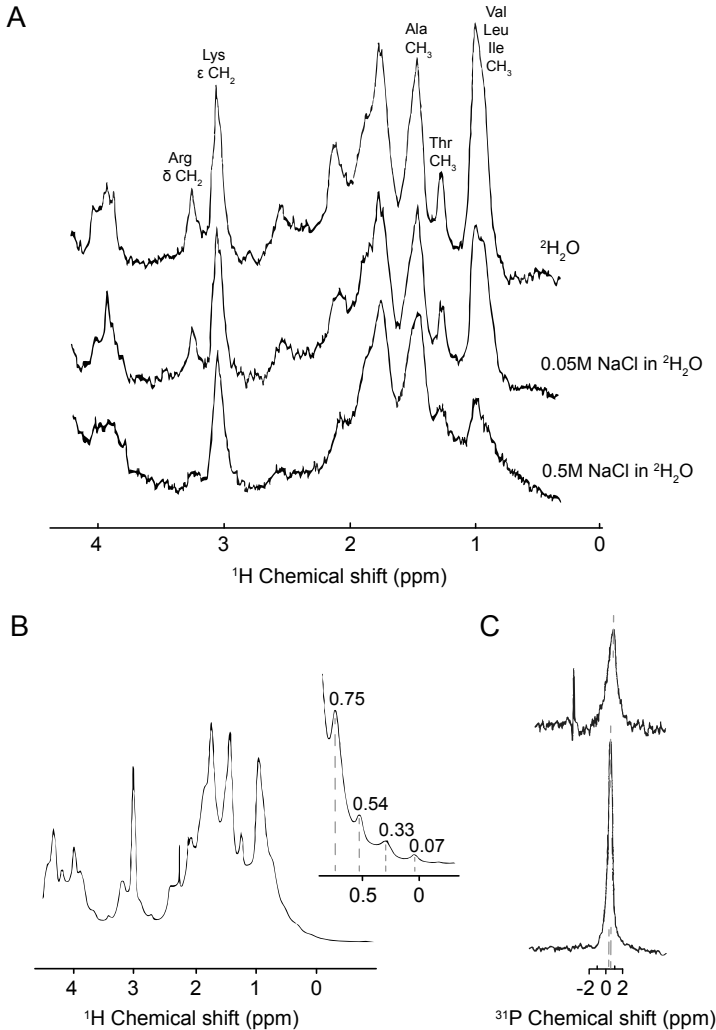


Figure 2. Pioneering NMR studies on nucleosomes and histones. (A) 1D ¹H NMR spectra (220 MHz) of histone H2B showing strongly reduced peak intensity for leucine, isoleucine and valine methyl group resonances at high ionic strength, while the lysine side chain resonance is unaffected. This suggested that leucine, isoleucine and valine experience restricted mobility upon salt-induced folding and interaction while lysines (present mostly in the N-terminal tail) remain mobile. Reproduced with permission from (45). (B) 1D ¹H NMR spectrum (270 MHz) of H2A-H2B dimers with the inset showing a deconvoluted, resolution-enhanced spectral expansion. The ring current shifted apolar resonances suggested that some methyl groups are in close contact with aromatic residues, indicating a specific interaction between H2A and H2B. Reproduced with permission from (58). (C) 1D ³¹P NMR spectra (measured at 60 MHz 1H

frequency) of nucleosome cores (top) and free nucleosomal DNA (bottom) showing that DNA in the nucleosome has a larger linewidth than free DNA but no significantly different chemical shifts, suggesting that DNA in the nucleosome is not kinked. Reproduced with permission from (71).

3.2. 'UNWRAPPING' THE NUCLEOSOMAL DNA

Acquiring and interpreting NMR spectra of nucleosomal DNA, corresponding to ~100 kDa of the 200 kDa nucleosome assembly, is extremely challenging due to extensive signal overlap and line broadening. Nevertheless, some bold attempts were made in the 1970's to assess the conformational state of DNA in chromatin. As it was known that DNA is present in 'beads' or v bodies in chromatin connected by linker DNA (72), the DNA inside the beads was assumed to adopt a conformation different from linker DNA, and several models were proposed for the wrapping of DNA around the histone core, either suggesting kinked bending every 10 or 20 bp (73,74), or continuous bending (75,76). Hanlon *et al.* (77) were the first to use ^{31}P NMR to probe the DNA conformation in purified, sheared chromatin fragments. They found that the severely broadened ^{31}P signal likely corresponds only to linker DNA between the beads, suggesting a rigid structure and strong interaction of the backbone phosphates with the histones inside the beads. Subsequent studies used MNase-digested chromatin to result in smaller nucleosome particles and thus better spectra. These did not find significant ^{31}P chemical shift perturbations (CSPs) compared to free DNA, suggesting smooth bending rather than kinking (Fig. 2C) (71,78-81). Scrutiny of upfield-shifted imino ^1H resonances suggested local alteration of base pairing near histone-DNA contacts (82,83), in agreement with the first crystal structure of the nucleosome (84). In 1997, the high-resolution nucleosome structure by Luger *et al.* (10) definitively confirmed that kinking occurs at several sites in the nucleosomal DNA, but not in the predicted regular way of every 10 or 20 bp.

Following up on the first ^{31}P ssNMR experiments on chromatin by DiVerdi *et al.* (81), the group of Kyogoku did some pioneering work in the late '80s on the structure of *in vivo* chromatin by probing the dynamic state of the DNA (85,86). Using ^1H - ^{31}P cross-polarization to distinguish the slow-tumbling chromatin from flexible phospholipids and inorganic phosphate, they measured the chemical shift anisotropy of ^{31}P in intact chicken erythrocytes, and concluded that the DNA is highly condensed, supporting the still highly debated presence of a 30 nm fiber.

Table 1. Milestones in structural studies of the nucleosome

Timeline	Reference	Key finding	Method
1884	Kossel (37)	Discovery of histone proteins	
1965	Phillips (38)	Five histone fractions	
1969-1972	DeLange, Iwai, Yeoman (40-43)	Histone primary sequences	
1974	Olins (72)	Beads on a string (ν bodies)	EM
1974-1975	Kornberg & Thomas (56,57)	Tetramers and dimers of histones, nucleosome composition	X-ray, cross-linking
1984	Richmond (84)	7 Å structure of the nucleosome	X-ray
1987	Clore (87)	Solution structure of histone H5	NMR
1993	Ramakrishnan (88)	Crystal structure of histone H5	X-ray
1994	Cerf (89)	Solution structure of histone H1	NMR
1997	Luger (10)	2.8 Å structure of the nucleosome	X-ray
1999	Dhalluin (90)	First titration of reader domain with acetyl-Lys histone peptide	NMR
2002	Nielsen, Jacobs (91,92)	First structure of reader domain in complex with methyl-Lys histone peptide	NMR, X-ray
2002	Davey (93)	1.9 Å structure of the nucleosome	X-ray
2005	Schalch (94)	Tetranucleosome structure	X-ray
2006	Barbera (95)	First structure of nucleosome-peptide complex (LANA)	X-ray
2010	Makde (96)	First structure of nucleosome-protein complex (RCC1)	X-ray
2011	Kato (97)	First high-resolution NMR study on the nucleosome and interaction study with (HMG2), methyl assignments	NMR
2013	Gao (98)	Tail dynamics in nucleosome arrays	ssNMR
2014	Song (99)	30 nm fiber structure	cryo-EM
2016	Moriwaki (100)	H2A-H2B dimer structure	NMR
2018	Xiang, Shi (101,102)	First high-resolution ssNMR study of nucleosome histone core	ssNMR

3.3. STRUCTURE OF THE LINKER HISTONE

The presence of linker histone H1 was established early on in chromatin research, and it was known to have a high lysine, alanine and proline content and less tendency to aggregate than the core histones. It is the first histone to dissociate from chromatin upon increase of ionic strength and already in the 1960's, it was clear that it can interact with naked DNA. In 1973, a 1D ^1H NMR study from Bradbury and colleagues reported histone H1 to be required for the condensation of chromatin and mostly bound to it by its lysine residues (103). In subsequent NMR papers, the focus has been mainly on identifying the position of the globular segment of this lysine-rich histone fraction (104-111). The high similarity between NMR spectra of H1 in chromatin and H1 bound to DNA indicated that H1 does not interact with the other histones (106), which was supported by neutron diffraction studies showing that H1 binds to the outside of a nucleosome (112).

Preceded by a preliminary study from the Bradbury lab using solely 1D spectra (113), in the mid 1980's Clore *et al.* succeeded in obtaining resonance assignments of an avian-specific isoform of the linker histone, histone H5. Using histones purified from chicken erythrocytes, they successfully applied the homonuclear NOESY-TOCSY approach developed by Wüthrich a few years earlier (114-116). As one of the first protein structures ever to be determined by NMR, a structural model of H5 was constructed based on 307 interproton distance restraints from NOE data elucidating the helical organization of the protein (87). Later work in the mid 1990's on the globular domain of the H1 isoform used a recombinant polypeptide and ^{15}N isotope labeling to determine the tertiary structure of the globular domain of linker histone H1 (89,117). The overall structure was found to be very similar to the H5 crystal structure (88), with small, local structural differences suggested to be responsible for the difference in DNA binding affinity between linker histones H1 and H5.

3.4. NON-HISTONE PROTEINS IN CHROMATIN

Apart from histone proteins, chromatin isolated from cells contains a small but significant fraction of non-histone proteins (118). Particularly abundant are the so-called high-mobility group (HMG) proteins (119). Proteins of the HMGN subclass bind the nucleosome, and this interaction was the first nucleosome-protein interaction to

ever be studied. Again, it was work from Bradbury from the late 70's that showed by 1D ^1H NMR that HMGN1 and HMGN2 are in a random coil conformation in their free state, contain a DNA-binding region between residues 15 and 40 and are completely bound to sonicated calf thymus DNA at low ionic strength (120,121). In 1989, Cook *et al.* concluded from ^1H NMR data that the basic central region of HMGN2 binds strongly to the nucleosome core particle, but they also detected weak binding of the acidic C-terminal region. They assumed that these basic residues would interact with the acidic backbone of the DNA, and the acidic C-terminal region presumably to the core histones (122). As will be described in section 6, it took the development of Methyl-TROSY (MeTROSY) NMR to demonstrate that the central basic region binds both the acidic patch on the surface of the H2A-H2B dimer and the DNA simultaneously (97).

3.5. SUMMARY

Well before the development of structure determination by NMR, multidimensional NMR and isotope labeling, NMR had already made an important contribution to the characterization of histone structure and interactions. In the pioneering phase of nucleosome research, before the 1984 structure (84), the composition and the overall shape of the nucleosome was known from X-ray diffraction and electron microscopy (56,57). During this stage, NMR was able to identify the position of the histone helices and ascertain that H2A-H2B dimers and H3-H4 tetramers have a defined, specific fold. NMR studies were instrumental in establishing the concept of histone tails: a large body of work showed that the N-terminal segments of the core histones are unstructured, highly flexible and weakly associated to the DNA. Finally, NMR was a crucial tool in determining the structure of the linker histones H1 and H5 and in characterizing the interaction of HMGNs with the nucleosome.

4. MODERN NMR STUDIES ON THE NUCLEOSOME

The success of modern-day NMR, as with any other structural biology technique, greatly depends on the ability to obtain homogeneous, stable samples, and this is no less true for studies of the nucleosome. Difficulties in reconstituting nucleosomes *in vitro* proved a great challenge for structural studies. The Herculean effort by Luger,

Richmond and co-workers to solve the first high resolution crystal structure of the nucleosome showed that a high affinity DNA sequence was crucial for efficient reconstitution of nucleosomes *in vitro* (10). Thanks to method development in solution and ssNMR, the challenge of the high molecular weight of the nucleosome was overcome, permitting the atomic-resolution structural and dynamical characterization of nucleosomes. In this section, we will detail the sample production and NMR methods available for such studies.

4.1. SAMPLE PREPARATION

Nucleosomes for NMR experiments are reconstituted from recombinantly expressed histones and nucleosomal DNA of a defined sequence. Histones are expressed separately, purified and refolded into octamers. Subsequent salt gradient-mediated deposition of octamers onto the DNA yields nucleosomes. A great advantage of this stepwise procedure is that it allows differential isotope labeling of the histones to reduce signal overlap.

4.1.1. PRODUCTION, ISOTOPE LABELING AND MODIFICATION OF HISTONES

Histones are extremely well-conserved proteins, giving the experimenter a certain degree of freedom in the choice of construct. Most NMR studies have used either *Drosophila* (Kay, Bai, Van Ingen labs), *Xenopus* (Narlikar lab) or human histones (Mer, Nishimura, Pervushin & Nordenskiöld labs). Histones are generally overexpressed from codon-optimized plasmids in *E. coli*, according to the protocol published by the Luger lab (123). As isolated histones are insoluble, they are subsequently purified from inclusion bodies, followed by gel filtration and ion exchange chromatography under denaturing conditions. Yields are typically around 10-50 mg per liter of culture, depending on the histone and culturing medium. Denatured, purified histones can be refolded into histone dimers, tetramers or octamers by dialysis to high salt buffer and purified by size exclusion chromatography with a typical 60-80% final yield. To optimize the yield and prevent aggregation, it is essential to ensure an equimolar ratio of the histones. Because histones carry few aromatic side chains and are devoid of tryptophan residues, determination of the protein concentration is particularly sensitive

to contamination with DNA. An additional ion-exchange chromatography step may be required to obtain DNA-free, pure histones (124). Histones are generally expressed separately but there are a few reports of successful use of H2A-H2B fusion proteins (125,126).

When using uniform $^{15}\text{N}/^{13}\text{C}$ isotope labeling, production in D_2O -based media with protonated glucose is sufficient to obtain high-quality spectra of histone dimers or the highly flexible tails within the nucleosomes. In addition, uniform labeling has been shown to be suitable for ssNMR studies of the nucleosome (see below) (101,102). To probe the rigid core of the 200 kDa nucleosome complex by solution NMR, the most suitable labeling strategy is the specific labeling of methyl groups in a highly deuterated background, as introduced by the Kay lab (127). Specific methyl group labeling is achieved by expressing the histone in fully deuterated M9 medium in the presence of amino acid precursors that are $^1\text{H}^{13}\text{C}$ -labeled only at the methyl group of interest. Protocols for labeling the isoleucine- δ 1, leucine and valine (ILV), or alanine, methionine and threonine methyl groups are reviewed in Kerfah *et al.* (128).

Several strategies can be used to incorporate PTMs into histone proteins (reviewed in (129)). However, their combination with isotope labeling remains challenging. Since genetic approaches relying on amber codons typically suffer from low yields, chemical approaches based on modification of an introduced cysteine residue have been used to install a methylated lysine mimic (125,130-132) and a ubiquitinated lysine (125,133). Histones contain only one conserved native cysteine, H3 C110, which can be substituted to Ala or Ser without loss of function.

4.1.2. PRODUCTION OF NUCLEOSOMAL DNA

DNA for the reconstitution of nucleosomes can be of varying length and sequence. Most commonly, a strong-positioning sequence is used that ensures efficient reconstitution of nucleosomes. These sequences, such as the non-natural '601' sequence from Widom (14) or the human α satellite repeat (10), are AT-rich in regions where the minor groove faces the octamer to ensure uniform positioning of the nucleosome. Minimum length for nucleosome core particle reconstitution is 145-147 bp; longer sequences will provide linker DNA extending from the core particle.

Nucleosomal DNA can be produced from a multiple-copy plasmid that is amplified in *E. coli* and extracted from the cells by alkaline lysis (123). The plasmids are

subsequently digested to separate the repeats and purified by ion exchange chromatography. This method typically yields about 20 mg of 147 bp 601-DNA from 3 liters of culture. Alternatively, nucleosomal DNA can be obtained through large-scale PCR followed by ion exchange chromatography. Although this method is labor-intensive and ideally requires several PCR machines running in parallel, 2-3 mg of 601-DNA can be produced from 4000 50 μ L PCR reactions in less time compared to the plasmid method. Plasmid-based production of DNA readily allows isotope labeling by culturing the plasmid in ^{13}C - or ^{15}N -labeled medium, although severe signal overlap due to the poor chemical shift dispersion of DNA can be expected.

4.1.3. RECONSTITUTION OF NUCLEOSOMES AND NUCLEOSOME ARRAYS

Nucleosomes are reconstituted by salt-gradient deposition, in which histone octamers and DNA are combined at 2M salt concentration and gradually dialyzed to low salt conditions to allow the nucleosome to form in a stepwise manner (123). Also here, accurate determination of octamer and DNA concentration is vital to ensure an equimolar mix and to avoid aggregation. Proper tuning of this ratio will also prevent excess free DNA in the final preparation, which can be determined from native PAGE analysis. Yields for reconstitution are typically around 80-90%. Sample conditions for solution NMR are usually 50-150 μ M nucleosomes in a 10-20 mM Tris or NaPi buffer at pH 6-8 and 0-150 mM NaCl or KCl. Experiments are usually recorded at 20-45 $^{\circ}\text{C}$.

Nucleosome arrays are reconstituted in a similar way by stepwise salt dialysis using multiple repeats of a strong-positioning sequence including linker DNA segments (134). Octamers and DNA are combined in the presence of lower affinity competitor DNA, to prevent overloading of the DNA with octamers. The arrays are purified by Mg^{2+} -precipitation or by sucrose gradient centrifugation (98). Increasing the divalent cation concentration in the preparation will promote array compaction and self-association.

4.2. STATE-OF-THE-ART NMR METHODS

A wide range of NMR methods is available to study the structural and dynamic properties of (sub)nucleosomal complexes. Depending on the question at hand, a divide-and-conquer strategy may be employed, allowing data to be recorded on histone peptides or histone dimers (~25 kDa) using the entire toolbox of protein solution NMR. In the sections below, we focus on the state-of-art methods that enable the study of intact nucleosomes or nucleosome-protein complexes by NMR, summarized in Figure 3.

histone tail	solution NMR	INEPT-based	NOESY PRE, PCS backbone chemical shifts	^{15}N T_1 , T_2 , NOE CPMG, CEST	CSPs saturation transfer H/D exchange intermolecular NOESY
	ssNMR	INEPT-based	PRE, PCS backbone chemical shifts		CSPs magnetization transfer
		strategy	structure	dynamics	interactions
histone core	solution NMR	MeTROSY-based	NOESY, RDC PRE, PCS	Me ^1H T_2 , S^2 $^{13}\text{C}/^1\text{H}$ MQ CPMG CEST	CSPs saturation transfer intermolecular NOESY
	ssNMR	CP-based	dipolar recoupling PRE, PCS backbone chemical shifts	dipolar couplings S^2 ^{15}N $T_{1\rho}$	CSPs (1H detection) magnetization transfer

Figure 3. NMR tool box for the study of nucleosome structure, dynamics and interactions.

4.2.1. METHYL-BASED SOLUTION NMR

The development of the MeTROSY approach has been instrumental in pushing the molecular size limitations of solution NMR into the MDa range (6,135). In this approach, side chain methyl groups, typically in isoleucine, leucine and valine, are used as local probes for structure and dynamics. Methyl groups offer three distinct advantages as NMR probes: being at the tip of flexible side chains, they are partially decoupled from the slow overall molecular tumbling, they contain three equivalent protons, and they rotate rapidly around their internal symmetry axis, allowing the exploitation of a line-narrowing effect caused by relaxation interference between different dipolar interactions within the methyl group (136). The HMQC experiment takes full advantage of this effect and is thus the cornerstone of MeTROSY, offering high-quality, sensitive spectra of large complexes. Importantly, the presence of external protons causes efficient relaxation of the methyl protons, especially for high-molecular-weight systems. Thus, near-complete deuteration of the non-methyl protons is essential to achieve the highest sensitivity. In the context of nucleosome studies, the DNA used for reconstitution is protonated and indeed histone methyl groups close to the DNA are significantly broadened (see also Fig. 4). Likewise, use of protonated binding partners in nucleosome-protein interaction studies will result in peak broadening for interface residues.

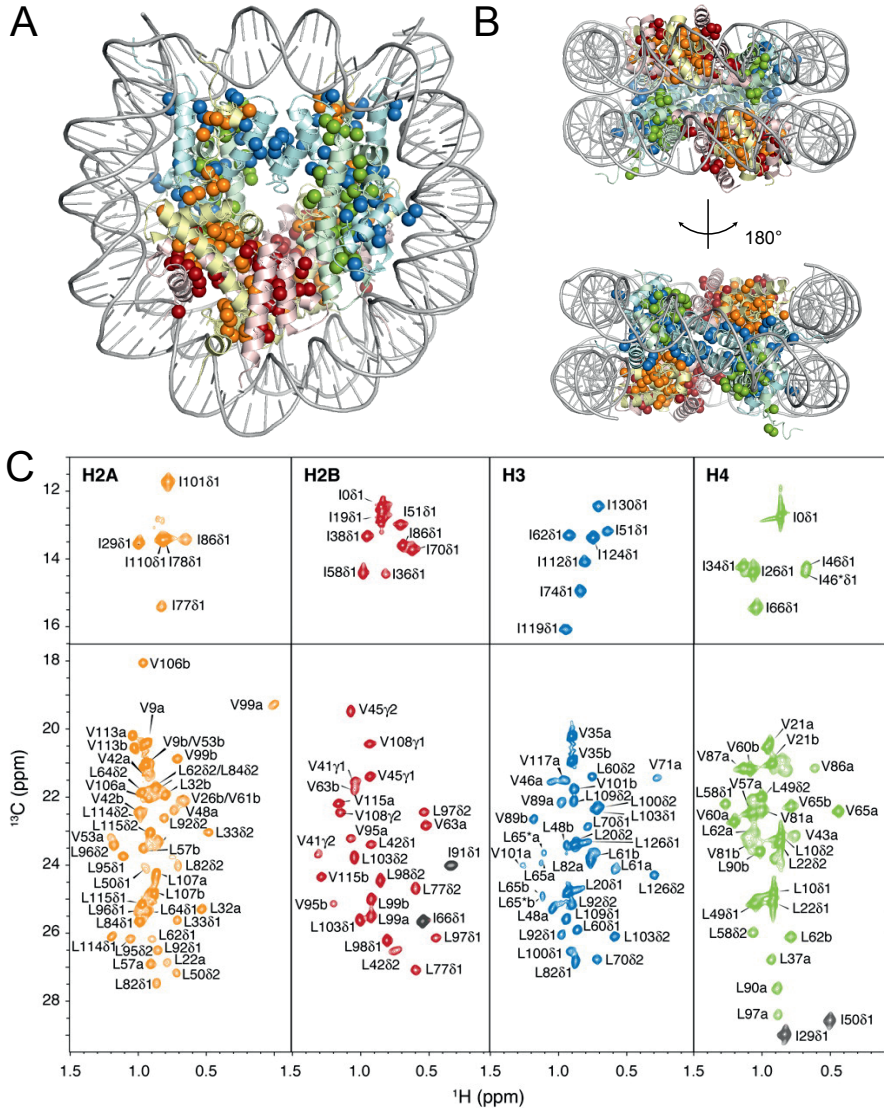


Figure 4. Histone methyl groups as molecular spies in the MeTROSY approach. Top (A) and side (B) views of the nucleosome (PDB entry 2PYO) showing the position of all ILV-methyl groups (spheres) in the core histones. (C) MeTROSY spectra of ILV-methyl labeled histones H2A, H2B, H3 and H4 in the nucleosome. Stereospecific assignments for leucine and valine are indicated by δ 1/ δ 2 and γ 1/ γ 2 where available, otherwise the two methyl groups are arbitrarily assigned as 'a' or 'b'. The resonances of L65 in histone H3 are significantly broadened due to their close proximity to the DNA, and occur in two sets (labeled L65 and L65*) because the local DNA sequence differs. Reproduced with permission from (97).

In a collaboration between the Kay and Bai labs, the MeTROSY approach was first applied to the nucleosome (97). The ILV-methyl groups in the nucleosome are reasonably distributed over the tail and core regions, including both the histone octamer surface and interior (Fig. 4A and 4B). MeTROSY spectra of the nucleosome in which one of the histones is ILV-labeled are of excellent quality (Fig. 4C). Assignment of the NMR signals is required for the site-specific interpretation of the structural and dynamical properties they encode. A near-complete assignment (89%) was achieved based on a combination of structure-based NOESY assignment and extensive mutagenesis, and the assigned chemical shifts are reported in the supplementary material of ref. (97). Using samples in which both methyl groups in Leu and Val residues are $^1\text{H}/^{13}\text{C}$ -labeled, the corresponding resonances were paired based on strong intra-residual cross-peaks in NOESY spectra recorded with short mixing time. To provide unambiguous check points for the assignment, in total 59 Ile to Leu or Leu/Val to Ile point mutations were made, out of 170 methyl groups in the nucleosome. The assignment was then completed by matching the NOE pattern in a set of six NOESY spectra, obtained on samples containing either ILV-labeled H2B, H3, H4, H2A/H2B, H2A/H3, or H3/H4, to the network of short methyl–methyl distances in the crystal structure. For several Leu/Val residues stereospecific assignments could be obtained.

The workhorse of the MeTROSY suite of experiments is the 2D $^{13}\text{C}^1\text{H}$ HMQC experiment. High quality 2D correlation maps can typically be obtained in ~ 1 hour on samples containing 100 μM nucleosomes (200 μM of each histone), corresponding to ~5 mg of nucleosomes in a 250 μL Shigemi tube sample. These can not only be used for CSP mapping, but also for the measurement of long-range distance restraints by exploiting either pseudocontact shifts (PCS) or paramagnetic relaxation enhancement (PREs) generated by a paramagnetic spin label (137). The HMQC pulse scheme is also amenable to implementation of very fast pulsing schemes that enable the study of real-time kinetics (138,139). In addition, MeTROSY-optimized pulse sequences are available for the measurement of residual dipolar couplings (RDCs) (140), NOE distance restraints, and the study of protein dynamics. Side-chain motions on a fast ps-ns timescale can be extracted from ^1H transverse relaxation rates and intra-methyl ^1H - ^1H dipolar cross-correlated relaxation rates (141), while slow μs -ms motions can be studied using $^{13}\text{C}^1\text{H}$ multiple-quantum Carr-Purcell-Meiboom-Gill (CPMG) relaxation

dispersion (142) or methyl-based chemical exchange saturation transfer (CEST) experiments (143,144).

4.2.2. SOLID-STATE NMR

As chromatin is a polymer that can easily be precipitated by the addition of divalent ions, it is a very suitable system to be investigated by ssNMR. Moreover, the breakthrough demonstration that sedimentation of large, soluble protein complexes can result in highly resolved ssNMR spectra (145,146) also opened the door to studies of mononucleosomes or nucleosome core particles by ssNMR.

The few pioneering studies that have applied ssNMR to nucleosomes will be discussed in more detail in the next section. Apart from the lack of intrinsic size limit, ssNMR offers three additional technical advantages. One is the ability to observe all ^{15}N and ^{13}C heteronuclei in the histone cores without being restricted to methyl-group containing residue types as in the MeTROSSY approach. The second advantage is that the reduced labeling requirements also mean reduced costs for sample preparation. A final advantage is that by using either cross-polarization or scalar coupling-based polarization transfer, spectra can be edited to be sensitive to either the rigid core or mobile tails of the nucleosome. On the other hand, the sensitivity and resolution of ssNMR spectra are generally lower than for MeTROSSY solution NMR and titration experiments are not feasible, limiting interaction studies to comparisons of free and bound spectra. Regions with intermediate dynamics appear neither in INEPT or CP-based spectra, and are thus not observable in ssNMR, akin to exchange-broadened residues in solution NMR.

To obtain samples suitable for ssNMR, nucleosomes need to be in a highly dense phase ($\sim 0.3\text{-}0.5$ g/L), not unlike the nuclear nucleosome concentration. The extreme molecular crowding and packing of nucleosomes in these samples prevents overall tumbling, thus allowing cross-polarization methods to work. Two recent studies, one from the Pervushin and Nordenskiöld groups (101) and one from Xiang and Paige *et al.* from the laboratory of the present authors (102), demonstrate two methods that allow such samples to be obtained. Shi *et al.* used high concentrations of Mg^{2+} to precipitate nucleosomes and nucleosome arrays into the NMR rotor. In the Xiang and Paige *et al.* study the highly dense phase was created by sedimentation of

nucleosomes under low Mg^{2+} conditions by overnight ultra-centrifugation (102). In both cases, highly resolved spectra were obtained, allowing access to structural and dynamical information for nearly all histone core residues.

These two studies also illustrate two different overall approaches. The Pervushin and Nordenskiöld work relied on ^{13}C -detected ssNMR and recorded spectra from 3.2 mm rotors filled with 46 mg of nucleosomes or 1.9 mm rotors containing 15 mg nucleosomal arrays. Recording of 2D ^{13}C - ^{13}C correlation maps typically takes overnight under these conditions. The other work relied on 1H -detection, which necessitates fractional deuteration of the isotope labeled histone and the use of a high-speed magic angle spinning (MAS) set-up. Rotors with a 1.3 mm diameter that can be spun at 50-60 kHz MAS were filled through sedimentation using a custom-made filling device with ~2 mg of nucleosomes. Measurement of a 2D 1H -detected NH spectrum typically then requires overnight acquisition, illustrating the sensitivity gain over ^{13}C -detection.

In both strategies, assignment of histone resonances follows well-established ssNMR procedures. For the ^{13}C -detected approach this is based on 2D and 3D NCA, NCO, NCACX, NCOCX and CANCO experiments. For the 1H -detected strategy, triple-resonance 3D CANH, CONH, CA(CO)NH and CO(CA)NH experiments are used. While the Nordenskiöld and Pervushin groups focused their work on histone H4, Xiang and Paige *et al.* focused on histone H2A. In both cases, the histone of interest was $^{13}C/^{15}N$ -labeled, with all other histones and DNA left unlabeled. Histone H2A was also fractionally deuterated to remove 1H - 1H dipolar couplings, in particular between the H_N and H_α , to enable 1H detection. In both cases near-complete backbone assignment of the histone core was obtained, and the native histone fold could be confirmed using the assigned chemical shifts.

Internal dynamics of the histones were probed in two different approaches. For H2A, a mostly qualitative assessment of backbone mobility was carried out, simply using the assigned chemical shifts and editing the spectra using either INEPT to observe the histone tails or CP-based transfer to probe the histone core. In a rigorous approach, Shi *et al.* used 3D DIPSHIFT experiments (147) to derive the order-parameter S^2 for both the C_α - H_α and N - H_N bond vectors for each residue of H4 experimentally.

4.3. INTEGRATIVE MODELING

Structural data obtained by NMR on large macromolecular assemblies such as the nucleosome typically represent either an ambiguous list of residues likely to form an intermolecular interface or sparse intermolecular distance restraints. To translate such structural information into the 3D structures of nucleosome-protein complexes, computational docking approaches, for instance using HADDOCK (148,149) or IMP (150) are needed that integrate the NMR data with other interaction data such as those obtained from mutagenesis, cross-linking or small-angle X-ray scattering (SAXS). Clearly, these modeling approaches require knowledge of the 3D structures of both interaction partners and perform best when conformational changes upon binding are minimal. For a detailed overview of the strategies used in integrative modeling based on NMR data we refer the reader to specialized reviews (151-153).

5. UNSPINNING CHROMATIN IN HIGH-RESOLUTION

Modern sample preparation and NMR techniques have been applied to answer a wide range of questions on the structural conformation and dynamic properties of nucleosomes. In this section, we will review studies on the conformation of the nucleosomal DNA, chaperone-assisted nucleosome assembly, the linker histone and the histone tails. We will discuss several NMR strategies that have been developed to study the real-time deposition of histone tail modifications and the dynamics of the nucleosome core.

5.1. HIGH-RESOLUTION NUCLEOSOMAL DNA STUDIES

The pioneering NMR studies on nucleosomal DNA described in section 3 lacked the resolution and sensitivity to give detailed information on its conformation. Recently, Xu *et al.* applied a divide-and-conquer approach to relate DNA sequence to backbone conformation and nucleosome affinity, by cutting 601-DNA into 12 bp fragments and studying these in 2D $^1\text{H}^{31}\text{P}$ correlation experiments (154). Using the phosphorus chemical shift as a reporter, they conclude that the sequence-dependent phosphate backbone conformation determines the minor groove width and that the minor groove width profile of 601-DNA fragments is highly similar to the variations in minor groove

width observed in the NCP crystal structures. In a follow-up paper they extended this method with internucleotide distances from NOE connectivities and RDCs, which allowed the characterization of additional internucleotide parameters like roll, twist and slide to further determine the intrinsic properties of nucleosomal DNA fragments in solution (155). Such high-resolution studies of the DNA in the context of nucleosomes is as yet beyond the current state-of-the-art.

5.2. CHAPERONE-ASSISTED NUCLEOSOME ASSEMBLY

Nucleosomes are assembled by deposition of H3-H4 tetramers and H2A-H2B dimers onto the nucleosomal DNA. *In vivo*, this process is guided by histone chaperones to prevent the formation of non-native complexes (156). In addition, these chaperones play an important role in the incorporation of histone variants. In order to understand the process of nucleosome assembly, several NMR studies have focused on the structural and dynamic properties of these histone complexes and their interactions with chaperones.

The smallest stable subsystem within the nucleosome is the H2A-H2B dimer (25 kDa). The solution structure of isolated H2A-H2B dimers, obtained from backbone chemical shifts using CS-Rosetta (157), shows that, while the core region is well-folded, several well-defined regions of the dimer in the nucleosome are disordered in the isolated dimer (Fig. 5A) (100). Presumably, these regions become rigidified and folded upon interaction with H3-H4 or the DNA in the context of the nucleosome. Interestingly, H/D and solvent exchange and $^1\text{H}^{15}\text{N}$ -NOE measurements indicate that these regions do retain some residual structure. Importantly, regions of the histone core that are recognized by chromatin factors and histone chaperones are relatively well-defined, including the positively charged DNA binding surface and a hydrophobic pocket in the H2B α 1- α 2 loop that are recognized by several histone chaperones.

Histone chaperones show a wide range of histone binding modes ranging from interactions between folded protein domains to purely disordered interactions. NMR is suitable for deciphering all these interaction types, and in particular those that involve disorder. For studies on folded chaperones, traditional structure determination and interaction studies are very suitable, as for instance in the study of H3 chaperone Asf1a (158). A nice illustration of using H/D exchange by NMR to determine the histone

binding interface is offered by Su *et al.* (159) for histone chaperone Vsp75. Here, NMR spectra offer atomic resolution mapping of increased protection from deuterium exchange in the histone-chaperone complex. Another class of chaperones contains intrinsically disordered regions that become folded only upon binding. Here, NMR has been used for the structure determination of several chaperone-histone variant complexes, using a chaperone-variant chimera in order to obtain a stable complex and allow a traditional NOE-based structure calculation (126,160,161). Many chaperones contain stretches of acidic residues. The importance of such regions for the recruitment and binding of histones was recently demonstrated by NMR for nucleoplasmin (162). Corbeski *et al.* showed from a series of NMR interaction studies that sparse aromatic residues in such acidic domains can be critical to obtain the functional, specific binding mode (163). At the other extreme is the spectacular case of histone H1 chaperone ProT α (164). This intrinsically disordered chaperone interacts with the disordered tail of H1 with picomolar affinity. Remarkably, NMR studies revealed that both proteins fully maintain their flexibility and disorder in the complex. The large opposite net charge and extended interaction surfaces of the two proteins allow an extremely tight interaction without requiring a defined, structured binding mode.

While most studies have focussed on structures of chaperone-histone complexes, NMR can also be used to determine the interaction dynamics quantitatively. This was nicely illustrated for the Chz1 chaperone using CPMG relaxation-dispersion experiments to determine the dissociation rate of the complex (160). Similar studies on chaperone-histone complexes in the presence of DNA or histone-DNA complexes may offer detailed insights into the nucleosome assembly process.

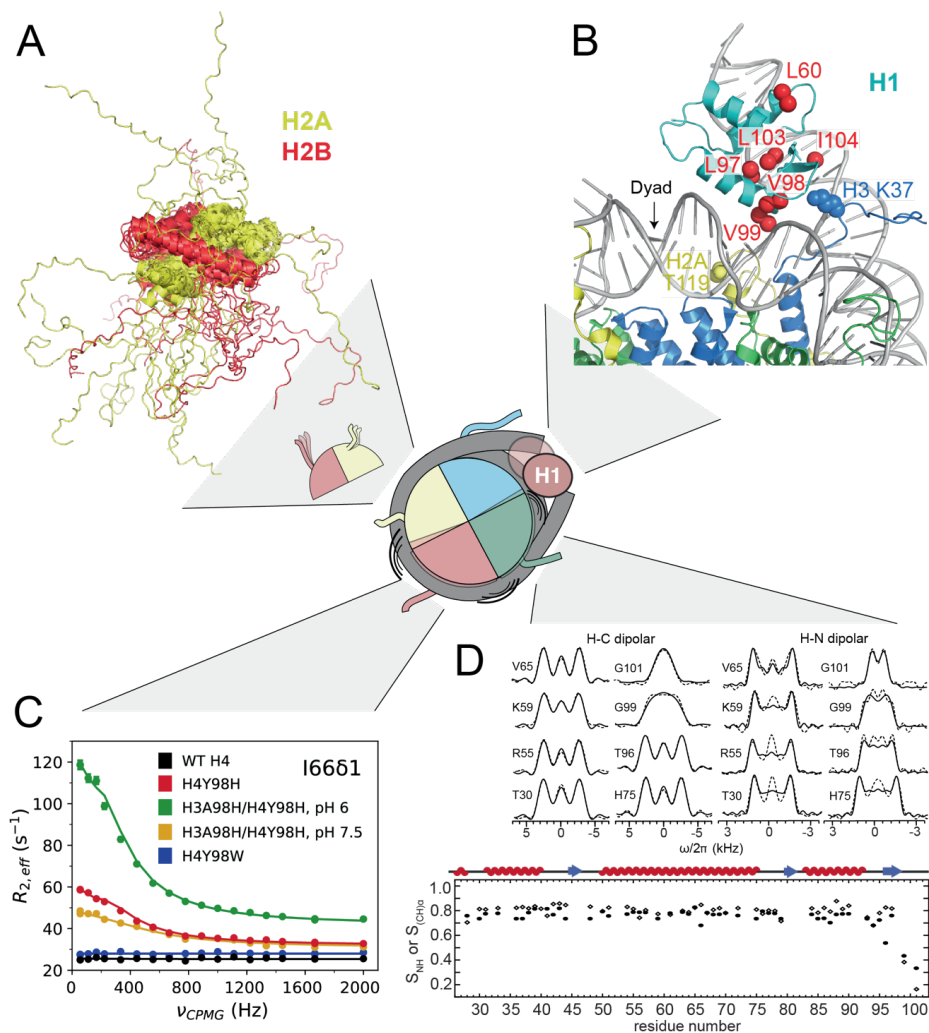


Figure 5. Schematic overview of NMR studies on histone structure and dynamics in the nucleosome. (A) Solution structure of the H2A-H2B dimer obtained by NMR (ensemble of 10 best scoring solutions based on CS-Rosetta, PDB entry 2RVQ), showing a native core histone fold as well as extensive unstructured tail regions. (B) NMR structure of the nucleosome based on MeTROSY PRE effects from spin-labeled H3K37 (blue spheres) or spin-labeled H2AT119 (yellow spheres) on ILV methyl groups in H1 (red spheres). Figure adapted with permission from (165). (C) MeTROSY CPMG relaxation dispersion profiles of ILV-labeled histone H2B (recorded at 600 and 800 MHz) show increasing slow timescale dynamics upon destabilizing mutations in H3 and H4. Figure reproduced with permission from (166). (D) Histone H4 dynamics obtained by fitting of ^1H - $^{13}\text{C}\alpha$ and ^1H - ^{15}N dipolar line shapes from 3D DIPSHIFT ssNMR experiments at 800 MHz (top) to derive $S_{(\text{CH})\alpha}$ and S_{NH} order parameters

(bottom). Figure reproduced with permission from (101). Color coding of histones and DNA as in Fig. 1.

5.3. THE CHROMATOSOME: ORIENTATION OF HISTONE H1

Recent studies on the structure of the chromatosome revealed the orientation and binding mode of histone H1 on the nucleosomal DNA. In 2013, the Bai lab reported an NMR-based structural model of the complex in which the globular domain of *Drosophila* H1 bridges the nucleosome core and the linker DNA by binding close to, but just off-center from, the central base pair at the dyad (165). Using the MeTROSY approach, they performed paramagnetic relaxation enhancement (PRE) experiments to obtain long-range distance information on the location and orientation of H1 and model its position on the nucleosome using computational docking (Fig. 5B). Such an off-dyad binding mode of the human linker histone is also reported by Song *et al.* (99), who solved the 11 Å structure of a 30 nm chromatin fiber by cryo-electron microscopy (cryo-EM). Interestingly, the first near-atomic resolution crystal structure of the chromatosome as well as a recent cryo-EM study showed on-dyad binding of chicken histone H5 and *Xenopus* histone H1, respectively (167,168). NMR and spin-labeling experiments revealed that five mutations in the globular domain of H5 into the corresponding residues in H1 can change the binding mode from on-dyad to off-dyad (169). The linker histone binding mode may be crucial in determining the conformation of higher order chromatin structure, which is emphasized by a recent study showing that the disordered H1 C-terminal tail promotes DNA compaction depending on its phosphorylation state (170).

5.4. MOBILITY OF HISTONE TAILS IN A CHROMATIN CONTEXT

In the past ten years, several sophisticated NMR studies were published reporting on the structure of the histone tails in the context of the full nucleosome and nucleosome arrays. Using MeTROSY and PRE experiments, the Bai lab showed that the basic patch in the tail of histone H4 folds onto the nucleosome core and that this interaction can be disrupted by acetylation of lysine 16 (171). The structure of the H3 tail in condensed nucleosomal arrays was addressed by the Bai lab using an H/D exchange experiment. After subjecting the arrays to H/D exchange, the histones are extracted in

DMSO to quench the reaction and dissolve the aggregates, before $^1\text{H}^{15}\text{N}$ HSQC detection of the backbone amide signals of H3 (172). Several tail residues of H3 showed slower exchange than expected for a purely unfolded state, suggesting that the H3 tail forms stable structures in condensed chromatin. Later, the Jaroniec lab used ssNMR to probe the structural and dynamic properties of the H3 and H4 histone tails in nucleosome arrays at varying degrees of Mg^{2+} -induced compaction (98). In contrast to the earlier report, they find that both the H3 and the H4 tail are highly dynamic in nucleosome arrays as they are observed in INEPT-based ^{13}C -detected ^1H - ^{13}C correlation spectra. The apparent difference from the Bai study may in part be due to the different readout chosen. In addition, the authors suggest that the histone tails could remain partly protected from solvent exchange while highly flexible. An insightful third study from the Selenko and Fischle labs added another dimension to the question of histone tail conformation. A series of solution NMR experiments, including ^{15}N -based relaxation measurements, showed that H3 tail flexibility in nucleosomes decreases upon inclusion of linker DNA and linker histone, which can be counteracted by introducing charge-modulating PTMs (173). It was also shown that the H3 tail has intrinsic DNA binding affinity, indicating that the H3 tail transiently interacts with nucleosomal (linker) DNA. These transient, PTM-dependent interactions may not only stabilize higher order chromatin, but also regulate accessibility to protein binding (see below). Recent ssNMR work showed that the H2A N-terminal tail has different chemical shifts in the nucleosomal context compared to the dimer, indicating it may be similarly bound to the DNA (102). Furthermore, peak splitting for several resonances matched the asymmetric DNA sequence close to the H2A tail location. The importance of PTMs for chromatin structure is further illustrated by a recent study on the effect of histone ubiquitination. This study employed H/D exchange in combination with NMR to map the interaction surface of H2BK120-conjugated ubiquitin in the context of 12-mer nucleosome arrays (174). A small acidic patch on ubiquitin was found to be involved in chromatin decompaction by acting as a dynamic wedge between nucleosomes in the array.

5.5. FOLLOWING NUCLEOSOME MODIFICATION IN REAL-TIME

Apart from structural and dynamics studies, NMR can also be used to study the modification of histones in real time. For example, binding affinities and conversion

rates for proline isomerization by the PPIase domain of Frp4 were determined by NMR for several H3 peptides by measuring the intensities of the exchange cross-peaks in $^1\text{H}^1\text{H}$ NOESY experiments (175). In an interesting approach, the Selenko and Schwarzer labs determined the activity of deacetylases and acetyl transferases on H4 tail peptides in HeLa nuclear extracts (176). Incorporation of NMR-active isotopes at selected lysine backbone and sidechain positions permitted the *in situ* observation of multiple acetylation events in parallel, in a time-resolved and quantitative manner. A subsequent study by the same lab used an elegant isotope labeling and histone purification scheme to study asymmetric modification of histones in the context of the nucleosome (177). Exploiting orthogonal, cleavable affinity tags, asymmetric H3-H4 tetramers were obtained in which one H3 chain is ^{15}N -labeled and the other ^{13}C -labeled. With this approach, PTM crosstalk in *cis* (within one H3 tail) and in *trans* (between two H3 tails in the same nucleosome) was analyzed by monitoring either a ^{13}C - or a ^{15}N -edited spectrum during enzymatic deposition of a second PTM.

5.6. NUCLEOSOME DYNAMICS

The nucleosome is subject to several dynamic processes, including assembly/disassembly, incorporation of histone variants and nucleosome remodeling, that are likely to require a certain amount of nucleosome deformation and plasticity. While less apparent than the flexibility of the histone tails, the dynamics of the nucleosome core is of particular interest to understand these processes. Two recent solution NMR studies show that the nucleosome core is indeed inherently dynamic and that this property is essential for successful remodeling. The Narlikar lab used MeTROSY NMR using nucleosomes containing ILV-labeled H4 to probe deformation of the histone core upon remodeling by SNF2h (178). They found that binding of SNF2h induces line broadening for several buried methyl groups, implying a conformational change of the histone core. This change could be mapped to ILV residues in the H3-H4 interface. Constraining this interface by site-specific cross-linking strongly inhibited nucleosome remodeling by SNF2h, suggesting that plasticity of the H3-H4 dimer interface is required for function. Similarly, a study from the Luger and Kay labs showed by MeTROSY CPMG relaxation dispersion measurements that destabilizing mutations in the dimer-tetramer interface between H2A-H2B and H3-H4 lead to significant (~8%) populations of 'excited' conformational states (Fig. 5C) (166).

While the ground state structure is unaffected by the mutation, the presence of these higher-energy conformations indicates that the H2A-H2B dimer can sample a more loosely associated state on a millisecond timescale. Such increased plasticity of the nucleosome could also be exploited by histone variants or PTMs to modify nucleosome function.

These studies are complemented by two recent ssNMR studies reporting on histone dynamics in the nucleosome core. The first study, by Xiang and Paige *et al.* from the present authors' laboratory, used ^1H -detected ssNMR to probe the structure and dynamics of H2A in sedimented nucleosomes (102). Based on observation of resonances in either INEPT or CP-based spectra together with backbone chemical shifts values, a clear trend in dynamics was observed: from highly flexible N- and C-terminal tails observable in the scalar-based spectra, to a rigid histone fold core. Notably, part of the αN helix and the C-terminal docking domain showed increased flexibility. In an elegant and thorough study, the Pervushin and Nordenskiöld labs determined the dynamics of histone H4 in the nucleosome to that in a 12-mer nucleosome array (101). By careful measurement and fitting of the $^1\text{H}_\text{N}$ - ^{15}N and $^1\text{H}_\alpha$ - $^{13}\text{C}_\alpha$ dipolar line shapes in N-CA correlation maps, order parameters and thus the extent of μs - ms backbone motions of H4 could be determined (Fig. 5D). The most dynamic regions in nucleosome and nucleosome arrays include residues involved in histone-DNA contacts, which could mean that these motions are important for DNA accessibility and nucleosome mobility.

6. NUCLEOSOME-PROTEIN INTERACTIONS

One of the defining features of the nucleosome is its function as a landing and binding platform for a wide array of proteins that control chromatin biology (see section 2). Structurally these interactions can be classified according to their interaction site on the nucleosome: the nucleosomal DNA, the histone tail or the nucleosome core surface (179). Nucleosome-binding proteins show a large variety of binding sites and modes, including low affinity or multivalent binding, and can be mediated through unstructured regions, relying predominantly on electrostatic interactions. Each of these interaction

types poses their own requirements and challenges when addressing them by NMR, and these are reviewed in this section.

6.1. STRUCTURAL BASIS OF THE HISTONE CODE

Since histone tails are generally considered to behave as independent units due to their flexibility, protein interactions with the tails have primarily been studied in a divide-and-conquer approach using histone peptides. Importantly, this allowed for easy implementation of PTMs such as lysine acetylation or methylation at defined positions through the use of solid phase peptide synthesis, as well as the use of relatively straightforward NMR methods to study the interactions of these peptides with protein domains.

In the earliest demonstration of specific readout of a histone modification, solution NMR was used to determine the first structure of a bromodomain. Titration experiments showed that this domain binds specifically to histone H4 tails carrying an acetylated lysine, and allowed mapping the location of the binding cleft onto the protein surface (90). Similarly, the specific readout of trimethylated lysine was first demonstrated using NMR titration experiments (180). The structure of the reader domain bound to a methyllysine histone peptide, determined both by NMR (92) and X-ray crystallography (91), showed that the recognition relies on the interaction of the trimethyl group with a cavity containing aromatic residues, coined the aromatic cage.

These NMR-based interaction studies provided crucial evidence and support for the so-called histone code hypothesis, as they revealed the structural basis and mechanism for its read-out (181-183). Such studies have also been instrumental in establishing the structural basis for the combinatorial readout of histone modifications, first demonstrated in the Allis lab by identifying a K9 methylation/S10 phosphorylation switch for the binding of heterochromatin protein HP1 to the H3 tail (184). This multivalent recognition of histone modifications can also be mediated by tandem reader domains, in which both domains each bind to a specific histone modification (185,186).

A similar set-up of NMR titrations of an isotope-labeled reader domain with modified histone peptides, often together with NMR solution structures of free and complexed

proteins, has been used extensively since. Two recent studies have extended this by using a spin-labeled or ¹⁹F-labeled version of the reader domain to characterize the histone interaction (187,188).

An as yet scantily covered area is the protein interactome of linker histone H1. The Thomas lab showed by NMR that the acidic tail of HMGB1, a member of the class B high mobility group proteins, interacts with the basic tail of H1, which may thereby promote the replacement of H1 with HMGB1 at distinct nucleosomal sites (189).

6.2. HISTONE TAIL INTERACTIONS IN THE NUCLEOSOMAL CONTEXT

The assumption of histone tails as completely independent units ignores the nucleosomal context in which they occur. Thus, the peptide-based studies described above potentially miss out on additional contributions of nucleosomal DNA or histone cores to the interaction, or the competing effect of the histone tail–DNA interaction that may reduce the availability of binding competent states (see section 5.4.).

A study from the Zweckstetter lab was the first to address this issue focusing on the interaction of heterochromatin protein HP1 with nucleosomes carrying the H3K9me3 modification (132). Comparing the interaction with modified nucleosomes and modified peptides, they found that the CSPs in HP1 were smaller in the nucleosome case, indicating that the nucleosomal context reduces the binding affinity two-fold. Importantly, the same set of HP1 residues showed CSPs and the perturbations were in the same spectral direction, evidencing that the binding mode is maintained between peptide and nucleosome. Notably, being bound to the histone tail, HP1 remains rather mobile in the complex, which is clear from the successful observation of the protein by amide backbone TROSY and further proven by relaxation experiments. Interestingly, a second ‘explorative’ binding mode was uncovered using unmodified nucleosomes in which HP1 weakly interacts with DNA.

Two recent studies shed more light on the origin of the inhibitory effect of the nucleosomal context. Gatchalian *et al.* investigated the interaction of nucleosomes with a paired reader domain, a construct containing two PHD domains that can each bind the H3 histone tail (190). Affinity and NMR measurements showed that nucleosome binding is impaired approximately six-fold compared to H3 peptide binding, while the

binding mode and the relative binding order of the PHD domains is maintained. Additional NMR experiments showed that one of the PHD domains in particular is repelled by the nucleosomal DNA. Combined with partial occlusion of the H3 tail due to DNA interaction, this likely accounts for the reduced affinity. Recently, Morrison *et al.* found a very pronounced inhibition of the interaction between the BPTF PHD finger reader domain and the H3K4me3 mark in the nucleosomal context (191). From a series of NMR titration experiments it was shown that the H3 tail transiently and dynamically interacts with nucleosomal DNA also in the nucleosome core particle, in the absence of any linker DNA. They further showed that the inhibition of tail-reader binding can be released by H3 modifications that weaken its interaction with the DNA (Fig. 6A).

The nucleosomal context can also promote histone tail interactions, as was shown for the recognition of the H3K36me mark. This modification site, K36, is close to the point where the H3 tail exits the core particle and thus any reader domain of H3K36me will be close to the nucleosomal DNA. This was first shown for the PSIP1 PWWP reader domain by Van Nuland *et al.* (130). NMR titration experiments showed that this reader domain binds with high millimolar affinity to a modified peptide and that this very weak interaction is completely dependent on the presence of the modification. In the context of modified nucleosomes however, a low micromolar affinity was found, corresponding to a dramatic increase in affinity of four orders-of-magnitude. Here, MeTROSY titration experiments were used to determine the microscopic affinity at the H3 tail site by fitting of the line shapes of perturbed H3 ILV residues. The structural model based on NMR and mutation data reveals how the PWWP reader domain is optimized to recognize the nucleosomal context of this modification by engaging simultaneously to the methylated histone tail and the DNA backbone.

In a study of Musselman *et al.* similar enhancement of affinity in the nucleosomal context was observed for the recognition of H3K36me by PHF1 Tudor domain (192). Interestingly, a high flexibility of a reader-nucleosome complex was observed, permitting the use of amide-observed TROSY NMR to map the binding surface on the reader domain. The structural model again showed that the reader domain can interact with histone tail and DNA simultaneously, rationalizing the enhanced affinity. This study also points out the potential impact of protein interactions on nucleosome organization or stability, since both NMR and FRET data revealed a highly dynamic

complex and partial DNA unwrapping. Protein binding may cause increased nucleosome opening and DNA breathing, opening up otherwise occluded binding interfaces on the DNA or histone octamer. This was also hypothesized by Richart *et al.*, who show that the chromoshadow domain (CSD) of HP1 α binds to histone H3 on a site that is located just within the nucleosome core (193).

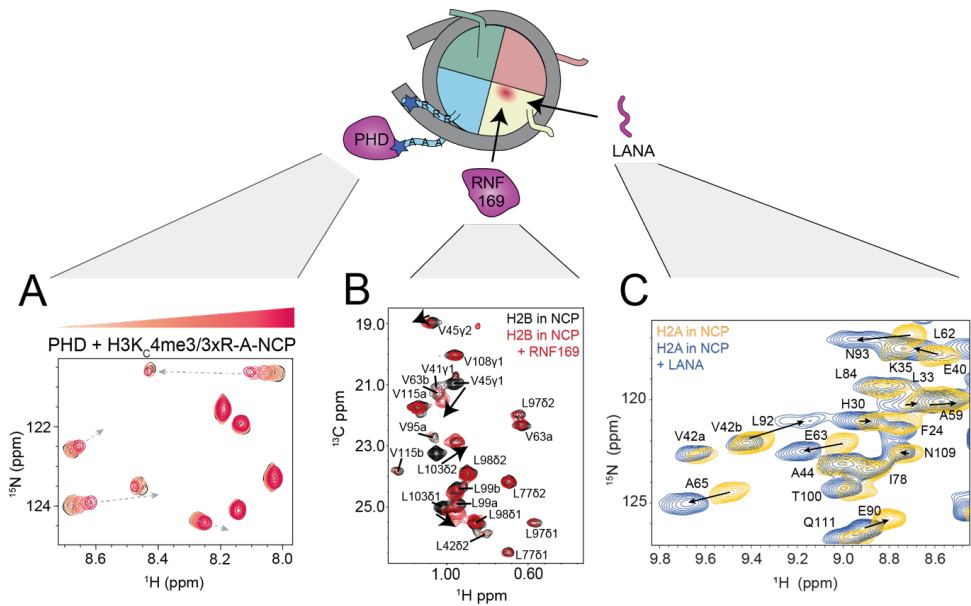


Figure 6. Schematic overview of nucleosome-protein interaction studies by NMR. (A) Recognition of the trimethylated K4 residue in the histone H3 tail (H3K4me₃) by the PHD reader domain as probed by NMR titration experiments. Addition of nucleosomes containing three arginine to alanine mutations in the H3 tail and a cysteine-based mimic for H3K3me₃ (H3K_C4me₃) result in clear and specific chemical shift changes for the reader domain. These mutations were necessary to weaken the tail-DNA interactions and increase the affinity of the nucleosome for the PHD reader domain. Figure reproduced with permission from (191). (B) Recruitment of RNF169 to nucleosomes ubiquitinated at H2A K13 or K15 depends on specific interaction of its UDM2 domain with the H2A-H2B acidic patch. Addition of this domain to ILV-methyl labeled H2B in H2AK13_CUb-NCPs resulted in several significant CSPs of acidic patch residues (indicated by arrows in the spectrum). Figure reproduced with permission from (133). (C) Mapping of the LANA-peptide binding interface on the nucleosome surface by ssNMR 2D NH spectra (800 MHz) on sedimented, free (yellow) and LANA-bound nucleosomes (blue) containing labeled H2A. Figure reproduced with permission from (102). Color-coding and symbols of the cartoon are as in Fig. 1, acidic patch is indicated in red.

6.3. DOCKING ONTO THE NUCLEOSOME CORE

In addition to histone tails, the histone core also features several sites for post-translational modifications that could act as a binding site for reader proteins (194-196). Many proteins have been identified over the last decade that bind to the nucleosome core. Apart from PTMs in the core, a negatively charged patch on the surface of the H2A-H2B dimer, often referred to simply as the acidic patch, has emerged as a prominent docking platform for many proteins that regulate chromatin function (24). Invariably, these proteins use an arginine residue to anchor to the acidic patch, without sequence motif or structural motif apart from hydrogen bonding. In favorable cases these interactions can be studied in a histone dimer context, but in general it requires the whole nucleosome to prevent missing out on detection of synergistic interactions with other histones or DNA. In addition, the opposite charge and different salt stability of the nucleosome and dimer may affect complex stability to a great extent.

The first study to show that NMR can resolve the structural basis of nucleosome-protein complexes was a collaboration between the Kay and Bai labs (97). This study pioneered the use of MeTROSY for nucleosomes and reported the assignments of the ILV methyl groups in all four histones (see section 4.2). Using these signals as probes, the interaction with the architectural chromatin factor HMGN2 was studied. This intrinsically disordered protein was known to anchor to nucleosomes but the exact binding mode and interaction surface on the nucleosome were unclear (see also section 3.4). On the basis of observed CSPs and saturation transfer experiments, the binding site of HMGN2 could be mapped to the acidic patch on the nucleosome surface. Additional mutational analysis revealed the crucial contribution of two conserved arginine residues to mediate the interaction. A series of PRE measurements using spin-labeled versions of HMGN2 revealed that its lysine-rich region is anchored to the nucleosomal DNA around the H3 tail exit site, highlighting the importance of studying such interactions in the nucleosomal context. The NMR and mutation data-driven structural model shows how this chromatin factor 'staples' nucleosomes through the two binding sites and thus orients its regulatory C-terminal domain to antagonize with linker histone binding, potentially destabilizing higher order chromatin.

Using the same MeTROSY approach, the Bai lab determined how centromeric protein CENP-C recognizes nucleosomes containing histone H3 variant CENP-A to initiate

assembly of the kinetochore during cell division (197). NMR experiments showed that a disordered segment of CENP-C is responsible for nucleosome binding. Through a combination of PRE data from spin-labeled versions of CENP-C and titration experiments, both short- and long-range contacts between CENP-C and histones in the nucleosome could be mapped. This pointed to a simultaneous engagement of CENP-C to a hydrophobic region in the CENP-A tail and the acidic patch of H2A and H2B, which was confirmed in a crystal structure (197).

A similar dual recognition mode was found for the recruitment of E3-ligase protein RNF169 binding to ubiquitylated nucleosomes in a study from the Kay lab (133). Here, the interaction depends on the ubiquitin-dependent recruitment module (UDM2) of RNF169, again containing a disordered region responsible for nucleosome binding. Ubiquitinated nucleosomes were prepared by converting G76 of ubiquitin and K13 of H2A to cysteine residues, conjugating them through sidechain-sidechain disulfide linkage before reconstitution. Nucleosomes contained either ILV-methyl labeled H2A, H2B or ubiquitin. MeTROSY-based relaxation measurements showed that the ubiquitin moiety is flexibly attached to the nucleosome. A series of titration experiments demonstrated that UDM2 binds to a hydrophobic patch on ubiquitin, and to the acidic patch of the nucleosome using two basic regions with arginine residues (Fig. 6B). Notably, a series of nucleosome and RNF169 mutants was screened by NMR to show not only the impact on binding affinity but also on binding mode, which proved crucial to dissect the different interactions to ubiquitin and the nucleosome core. The Mer lab studied the same interaction in a divide-and-conquer approach. First, the structure of a chimeric H2AK15Ub-H2B fusion protein in complex with UDM2 was determined based on intermolecular NOE and PRE restraints, together with backbone dihedral angle restraints (125). Together with additional SAXS data on the nucleosomal complex, the structure of nucleosome-bound RNF169 was modeled. Together, the studies described here illustrate that synergistic interactions are pervasively present in nucleosome-protein interactions.

The present authors' lab recently demonstrated the potential of ¹H-detected ssNMR in studying nucleosome-protein interactions (102). A key advantage here is the ability to observe all backbone NH correlation maps as in standard solution NMR, thus allowing the use of all non-proline amides as reporters on the interaction. ¹H detection is required to exploit the sensitivity of the ¹H spin to changes in the chemical environment

upon binding. Using the LANA-nucleosome interaction as a proof of principle, this study showed that it is possible to prepare a ssNMR sample of a complex through co-sedimentation of nucleosomes with the binding protein, in this case the LANA-peptide. Since titrations are not feasible in this set-up, CSPs have to be determined from a comparison of apo and bound state NH spectra. For unambiguous assignment of bound state resonances, a 3D HNCA experiment was recorded on the bound state. The changes in HNCA chemical shifts thus obtained accurately mapped the LANA binding site (Fig. 6C) and further confirmed that binding does not induce changes in H2A conformation from the lack of $C\alpha$ change, in agreement with the previously published crystal structure (95).

7. CONCLUSIONS AND OUTLOOK

Here, we have reviewed the contribution of NMR to nucleosome research, from the early work in the late 1960's pioneering structural investigations of histones and the conformation of nucleosomal DNA to the characterization of numerous reader-histone tail interactions and the most recent state-of-the-art NMR studies elucidating high-resolution structures of nucleosomes and their complexes with interacting proteins. Advances in both sample preparation and NMR methodology have gone hand-in-hand to deliver increasingly high-resolution information on nucleosome structure, dynamics and interactions. The development of MeTROSY-based solution NMR and advances in biomolecular ssNMR have been instrumental in pushing these studies to the level of the whole nucleosome.

A major open challenge for NMR is the assessment of DNA conformation and dynamics in the nucleosome. The molecular size and limited chemical shift dispersion of nucleosomal DNA essentially render the DNA invisible to either solution or solid-state NMR. Future experiments employing segmental or site-specific labeling (198) might overcome some of these challenges and allow direct and high-resolution detection of nucleosomal DNA. Alternatively, site-specific incorporation of paramagnetic tags into DNA (199) could enable indirect monitoring of dynamic changes in histone-DNA interactions. While it will require a formidable effort, such methods may open the door to future studies on the contribution of DNA in nucleosome-protein interactions, in particular in remodeling, the conformation and

flexibility of non-positioning, genomic DNA sequences, or the impact of DNA modifications.

Focusing on histones and nucleosome-binding proteins, we expect that on-going developments in NMR methodology will enhance the study of nucleosome-protein interactions in particular. Application of intermolecular PCS holds great promise to reveal quantitative long-range interaction data, complementing CSP-based identification of binding surface and sparse short-range NOE data. In addition, the sedimentation approach will be useful to characterize the structures of nucleosome-bound proteins and thus reveal conformational changes upon binding by ssNMR.

An intriguing question in the chromatin field concerns the impact of DNA sequence on nucleosome stability and dynamics. NMR studies on nucleosomes containing genomic DNA sequences will be highly instructive to expose the extent and timescales of histone dynamics in such nucleosomes, at atomic resolution. Moreover, the use of genomic DNA sequences also enables the incorporation of native binding sites for (pioneer) transcription factors to study their binding to the nucleosome (200).

It will be exciting to see how the study of these native-state interactions can be extended to higher order chromatin systems. The wide applicability of NMR to different sample phases, from dilute solution to condensed phases and cellular systems, allows investigation of biomolecular systems at different levels of complexity. We envisage that the combined use of solution, solid-state and in-cell NMR methods has the potential to enable multi-scale studies of chromatin systems and nucleosome-protein interactions, ranging from mononucleosomes to nucleosomal arrays to *in vivo* chromatin. By building on the intrinsic atomic resolution of the NMR signal and its sensitivity to molecular motions, NMR offers a unique perspective on the dynamic landscape of nucleosomes and their interacting proteins, which is crucial for our understanding of chromatin function.

8. SCOPE OF THIS THESIS

The studies presented in this thesis focus on the structural properties of the nucleosomal DNA and its role in interactions of the nucleosome with pioneer transcription factors and chromatin remodeler ISWI. New methods were introduced for the production of genomic nucleosomal DNA sequences and isotope-labeling of DNA to create new possibilities for structural studies on nucleosomal DNA by NMR.

Chapter 1 describes the contribution of NMR to our structural understanding of the nucleosome and nucleosome interactions. An extensive overview is provided of NMR studies ranging from the pioneering natural abundance NMR studies in the 1960's to state-of-the-art NMR experiments performed on nucleosome complexes today.

Chapter 2 investigates the architectural principles of histone complexes in archaeal chromatin using computational modeling. The models suggest that shape, symmetry and surface residues of the histone dimers determine the type of nucleosome-like complex that can be formed.

Chapter 3 introduces ramified rolling circle amplification (rRCA) as an efficient method for the large-scale production of nucleosomal DNA sequences for structural studies. rRCA allows for flexibility in the choice of sequence, which creates new possibilities to study pioneer transcription factor binding in nucleosomes reconstituted on native, genomic DNA.

Chapter 4 aims to contribute to our understanding of the mechanism of activation of chromatin remodeler ISWI. Fluorescence anisotropy and NMR data shed light on how the nucleosomal DNA and the histone H4 tail play a role in release of autoinhibition by ISWI's N-terminal region.

Chapter 5 explores the possibilities of isotope-labeling of nucleosomal DNA for NMR studies. The segmental, one-strand isotope-labeling of six thymine residues in DNA-601 provided local probes in the nucleosome that can be used to report on site-specific binding of interacting molecules, as shown for aclarubicin.

Chapter 6 provides a critical discussion of the presented studies and an outlook for further research on nucleosomal DNA by NMR.

REFERENCES

1. Lian, L.-Y. (2013) NMR studies of weak protein–protein interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **71**, 59-72.
2. Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S. and Allain, F.H.T. (2011) Structure determination and dynamics of protein–RNA complexes by NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **58**, 1-61.
3. Yadav, D.K. and Lukavsky, P.J. (2016) NMR solution structure determination of large RNA-protein complexes. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **97**, 57-81.
4. Zhuravleva, A. and Korzhnev, D.M. (2017) Protein folding by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **100**, 52-77.
5. Wang, S. and Ladizhansky, V. (2014) Recent advances in magic angle spinning solid state NMR of membrane proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **82**, 1-26.
6. Rosenzweig, R. and Kay, L.E. (2014) Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu Rev Biochem*, **83**, 291-315.
7. Kovermann, M., Rogne, P. and Wolf-Watz, M. (2016) Protein dynamics and function from solution state NMR spectroscopy. *Q Rev Biophys*, **49**, e6.
8. O'Connell, M.R., Gamsjaeger, R. and Mackay, J.P. (2009) The structural analysis of protein–protein interactions by NMR spectroscopy. *PROTEOMICS*, **9**, 5224-5232.
9. Campagne, S., Gervais, V. and Milon, A. (2011) Nuclear magnetic resonance analysis of protein–DNA interactions. *Journal of The Royal Society Interface*, **8**, 1065.
10. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.
11. Li, G. and Widom, J. (2004) Nucleosomes facilitate their own invasion. *Nature Structural & Molecular Biology*, **11**, 763.
12. Lilley, D.M.J., Howarth, O.W. and Clark, V.M. (1976) The existence of random coil N-terminal peptides — 'tails' — in native histone complexes. *FEBS Letters*, **62**, 7-10.
13. Fyodorov, D.V., Zhou, B.-R., Skoultchi, A.I. and Bai, Y. (2017) Emerging roles of linker histones in regulating chromatin structure and function. *Nature Reviews Molecular Cell Biology*, **19**, 192.
14. Lowary, P.T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*, **276**, 19-42.
15. Kugler, J.E., Deng, T. and Bustin, M. (2012) The HMGN family of chromatin-binding proteins: Dynamic modulators of epigenetic processes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1819**, 652-656.
16. Zaret, K.S. and Mango, S.E. (2016) Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current Opinion in Genetics & Development*, **37**, 76-81.
17. Bannister, A.J. and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Research*, **21**, 381.

18. DesJarlais, R. and Tummino, P.J. (2016) Role of Histone-Modifying Enzymes and Their Complexes in Regulation of Chromatin Biology. *Biochemistry*, **55**, 1584-1599.
19. Zhou, C.Y., Johnson, S.L., Gamarra, N.I. and Narlikar, G.J. (2016) Mechanisms of ATP-Dependent Chromatin Remodeling Motors. *Annual Review of Biophysics*, **45**, 153-181.
20. Clapier, C.R., Iwasa, J., Cairns, B.R. and Peterson, C.L. (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, **18**, 407.
21. Talbert, P.B. and Henikoff, S. (2016) Histone variants on the move: substrates for chromatin dynamics. *Nature Reviews Molecular Cell Biology*, **18**, 115.
22. Mattioli, F., D'Arcy, S. and Luger, K. (2015) The right place at the right time: chaperoning core histone variants. *EMBO Rep*, **16**, 1454-1466.
23. Fasci, D., van Ingen, H., Scheltema, R.A. and Heck, A.J.R. (2018) Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei. *Molecular & Cellular Proteomics*, **17**, 2018-2033.
24. McGinty, R.K. and Tan, S. (2015) Nucleosome structure and function. *Chem Rev*, **115**, 2255-2273.
25. McGinty, R.K. and Tan, S. (2016) Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol*, **37**, 54-61.
26. Rowley, M.J. and Corces, V.G. (2018) Organizational principles of 3D genome architecture. *Nat Rev Genet*, **19**, 789-800.
27. Li, G. and Zhu, P. (2015) Structure and organization of chromatin fiber in the nucleus. *FEBS Letters*, **589**, 2893-2904.
28. Sexton, T. and Cavalli, G. (2015) The role of chromosome domains in shaping the functional genome. *Cell*, **160**, 1049-1059.
29. Larson, A.G., Elnatan, D., Keenen, M.M., Tmka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S. and Narlikar, G.J. (2017) Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature*, **547**, 236.
30. Sewitz, S.A., Fahmi, Z. and Lipkow, K. (2017) Higher order assembly: folding the chromosome. *Current Opinion in Structural Biology*, **42**, 162-168.
31. Johns, E.W. (1964) Studies on histones. 7. Preparative methods for histone fractions from calf thymus. *Biochem J*, **92**, 55-59.
32. Murray, K., Vidali, G. and Neelin, J.M. (1968) The stepwise removal of histones from chicken erythrocyte nucleoprotein. *Biochem J*, **107**, 207-215.
33. Johns, E.W. (1967) A method for the selective extraction of histone fractions f2(a)1 and f2(a)2 from calf thymus deoxyribonucleoprotein at pH7. *Biochemical Journal*, **105**, 611-614.
34. Zubay, G. and Doty, P. (1959) The isolation and properties of deoxyribonucleoprotein particles containing single nucleic acid molecules. *Journal of Molecular Biology*, **1**, 1-20.
35. van der Westhuyzen, D.R. and von Holt, C. (1971) A new procedure for the isolation and fractionation of histones. *FEBS Letters*, **14**, 333-337.

36. Noll, M. and Kornberg, R.D. (1977) Action of micrococcal nuclease on chromatin and the location of histone H1. *J Mol Biol*, **109**, 393-404.
37. Kossel, A. (1884) Ueber einen peptonartigen Bestandtheil des Zellkerns. *Zeitschrift für physiologische Chemie*, **8**, 511-515.
38. Phillips, D.M. and Johns, E.W. (1965) A Fractionation of the Histones of Group F2a from Calf Thymus. *Biochem J*, **94**, 127-130.
39. Bradbury, E.M., Crane-Robinson, C., Goldman, H., Rattle, H.W.E. and Stephens, R.M. (1967) Spectroscopic studies of the conformations of histones and protamine. *Journal of Molecular Biology*, **29**, 507-523.
40. DeLange, R.J., Fambrough, D.M., Smith, E.L. and Bonner, J. (1969) Calf and pea histone IV. II. The complete amino acid sequence of calf thymus histone IV; presence of epsilon-N-acetyllysine. *J Biol Chem*, **244**, 319-334.
41. Iwai, K., Ishikawa, K. and Hayashi, H. (1970) Amino-acid sequence of slightly lysine-rich histone. *Nature*, **226**, 1056-1058.
42. DeLange, R.J., Hooper, J.A. and Smith, E.L. (1972) Complete amino-acid sequence of calf-thymus histone 3. *Proc Natl Acad Sci U S A*, **69**, 882-884.
43. Yeoman, L.C., Olson, M.O., Sugano, N., Jordan, J.J., Taylor, D.W., Starbuck, W.C. and Busch, H. (1972) Amino acid sequence of the center of the arginine-lysine-rich histone from calf thymus. The total sequence. *J Biol Chem*, **247**, 6018-6023.
44. Boublík, M., Bradbury, E.M. and Crane-Robinson, C. (1970) An Investigation of the Conformational Changes of Histones F1 and F2a1 by Proton Magnetic Resonance Spectroscopy. *European Journal of Biochemistry*, **14**, 486-497.
45. Boublík, M., Bradbury, E.M., Crane-Robinson, C. and Johns, E.W. (1970) An investigation of the conformational changes of histone F2b by high resolution nuclear magnetic resonance. *Eur J Biochem*, **17**, 151-159.
46. Bradbury, E.M., Cary, P.D., Crane-Robinson, C., Riches, P.L. and Johns, E.W. (1972) Nuclear-magnetic resonance and optical-spectroscopic studies of conformation and interactions in the cleaved halves of histone F2B. *Eur J Biochem*, **26**, 482-489.
47. Lewis, P.N., Bradbury, E.M. and Crane-Robinson, C. (1975) Ionic strength induced structure in histone H4 and its fragments. *Biochemistry*, **14**, 3391-3400.
48. Pekary, A.E., Chan, S.I., Hsu, C.J. and Wagner, T.E. (1975) Nuclear magnetic resonance studies on the solution conformation of histone IV fragments obtained by cyanogen bromide cleavage. *Biochemistry*, **14**, 1184-1189.
49. Pekary, A.E., Li, H.J., Chan, S.I., Hsu, C.J. and Wagner, T.E. (1975) Nuclear magnetic resonance studies of histone IV solution conformation. *Biochemistry*, **14**, 1177-1184.
50. Crane-Robinson, C., Hayashi, H., Cary, P.D., Briand, G., SautiÈRe, P., Krieger, D., Vidali, G., Lewis, P.N. and Tom-Kun, J. (1977) The Location of Secondary Structure in Histone H4. *European Journal of Biochemistry*, **79**, 535-548.
51. Bradbury, E.M. and Rattle, H.W.E. (1972) Simple Computer-Aided Approach for the Analyses of the Nuclear-Magnetic-Resonance Spectra of Histones. *European Journal of Biochemistry*, **27**, 270-281.
52. Clark, V.M., Lilley, D.M.J., Howarth, O.W., Richards, B.M. and Pardon, J.F. (1974) The structure and properties of histone F2a comprising the heterologous group F2a 1

- and F2a 2 studied by ^{13}C nuclear magnetic resonance. *Nucleic Acids Research*, **1**, 865-880.
53. Lilley, D.M.J., Howarth, O.W., Clark, V.M., Pardon, J.F. and Richards, B.M. (1975) Investigation of the conformational and self-aggregational processes of histones using hydrogen and carbon-13 nuclear magnetic resonance. *Biochemistry*, **14**, 4590-4600.
 54. Tancredi, T., Temussi, P.A., Paolillo, L., Trivellone, E. and Crane-Robinson, C. (1976) A study of calf-thymus histone H2B using ^{13}C magnetic resonance. *Eur J Biochem*, **70**, 403-408.
 55. Tancredi, T., Temussi, P.A., Di Pascale, G. and Fournier, C. (1979) Carbon magnetic resonance studies of the self-aggregation of calf thymus histones. *Eur J Biochem*, **100**, 219-224.
 56. Kornberg, R.D. and Thomas, J.O. (1974) Chromatin structure; oligomers of the histones. *Science*, **184**, 865-868.
 57. Thomas, J.O. and Kornberg, R.D. (1975) An octamer of histones in chromatin and free in solution. *Proc Natl Acad Sci U S A*, **72**, 2626-2630.
 58. Moss, T., Cary, P.D., Abercrombie, B.D., Crane-Robinson, C. and Bradbury, E.M. (1976) A pH-Dependent Interaction between Histones H2A and H2B Involving Secondary and Tertiary Folding. *European Journal of Biochemistry*, **71**, 337-350.
 59. Moss, T., Cary, P.D., Crane-Robinson, C. and Bradbury, E.M. (1976) Physical studies on the H3/H4 histone tetramer. *Biochemistry*, **15**, 2261-2267.
 60. Puigdomènech, P., Daban, J.R., Palau, J., Podo, F., Guidoni, L. and Temussi, P.A. (1977) The interaction of histone H3 with histone H4 and with other histones studied by ^{19}F nuclear magnetic resonance. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, **492**, 12-19.
 61. Böhm, L., Hayashi, H., Cary, P.D., Moss, T., Crane-Robinson, C. and Bradbury, E.M. (1977) Sites of histone/histone interaction in the H3 - H4 complex. *Eur J Biochem*, **77**, 487-493.
 62. Nicola, N.A., Fulmer, A.W., Schwartz, A.M. and Fasman, G.D. (1978) High resolution proton magnetic resonance spectroscopy of histones and histone-histone complexes in aqueous solution. *Biochemistry*, **17**, 1779-1785.
 63. Lilley, D.M.J., Pardon, J.F. and Richards, B.M. (1977) Structural investigations of chromatin core protein by nuclear magnetic resonance. *Biochemistry*, **16**, 2853-2860.
 64. Bradbury, E.M., Cary, P.D., Crane-Robinson, C., Rattle, H.W., Boublík, M. and Sautière, P. (1975) Conformations and interactions of histone H2A (F2A2, ALK). *Biochemistry*, **14**, 1876-1885.
 65. Cary, P.D., Moss, T. and Bradbury, E.M. (1978) High-Resolution Proton-Magnetic-Resonance Studies of Chromatin Core Particles. *European Journal of Biochemistry*, **89**, 475-482.
 66. Diaz, B.M. and Walker, I.O. (1983) Trypsin digestion of core chromatin. *Bioscience Reports*, **3**, 283.
 67. Cary, P.D., Crane-Robinson, C., Bradbury, E.M. and Dixon, G.H. (1982) Effect of Acetylation on the Binding of N-Terminal Peptides of Histone H4 to DNA. *European Journal of Biochemistry*, **127**, 137-144.

68. Walker, I.O. (1984) Differential dissociation of histone tails from core chromatin. *Biochemistry*, **23**, 5622-5628.
69. Smith, R.M. and Rill, R.L. (1989) Mobile histone tails in nucleosomes. Assignments of mobile segments and investigations of their role in chromatin folding. *Journal of Biological Chemistry*, **264**, 10574-10581.
70. Hilliard, P.R., Jr., Smith, R.M. and Rill, R.L. (1986) Natural abundance carbon-13 nuclear magnetic resonance studies of histone and DNA dynamics in nucleosome cores. *J Biol Chem*, **261**, 5992-5998.
71. Kallenbach, N.R., Appleby, D.W. and Bradley, C.H. (1978) 31P magnetic resonance of DNA in nucleosome core particles of chromatin. *Nature*, **272**, 134-138.
72. Olins, A.L. and Olins, D.E. (1974) Spheroid chromatin units (v bodies). *Science*, **183**, 330-332.
73. Sobell, H.M., Tsai, C.C., Gilbert, S.G., Jain, S.C. and Sakore, T.D. (1976) Organization of DNA in chromatin. *Proc Natl Acad Sci U S A*, **73**, 3068-3072.
74. Crick, F.H. and Klug, A. (1975) Kinky helix. *Nature*, **255**, 530-533.
75. Sussman, J.L. and Trifonov, E.N. (1978) Possibility of nonkinked packing of DNA in chromatin. *Proc Natl Acad Sci U S A*, **75**, 103-107.
76. Levitt, M. (1978) How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc Natl Acad Sci U S A*, **75**, 640-644.
77. Hanlon, S., Glonek, T. and Chan, A. (1976) Comparison of the phosphorus magnetic resonance and circular dichroism properties of calf thymus DNA and chromatin. *Biochemistry*, **15**, 3869-3875.
78. Cotter, R.I. and Lilley, D.M.J. (1977) The conformation of DNA and protein within chromatin subunits. *FEBS Letters*, **82**, 63-68.
79. Klevan, L., Armitage, I.M. and Crothers, D.M. (1979) 31P NMR studies of the solution structure and dynamics of nucleosomes and DNA. *Nucleic acids research*, **6**, 1607-1616.
80. Shindo, H., McGhee, J.D. and Cohen, J.S. (1980) Phosphorus-31 NMR studies of DNA in nucleosome core particles. *Biopolymers*, **19**, 523-537.
81. DiVerdi, J.A., Opella, S.J., Ma, R.I., Kallenbach, N.R. and Seeman, N.C. (1981) 31P NMR of DNA in eukaryotic chromosomal complexes. *Biochemical and biophysical research communications*, **102**, 885-890.
82. Feigon, J. and Kearns, D.R. (1979) 1H NMR investigation of the conformational states of DNA in nucleosome core particles. *Nucleic Acids Res*, **6**, 2327-2337.
83. McMurray, C.T., Van Holde, K.E., Jones, R.L. and Wilson, W.D. (1985) Proton NMR investigation of the nucleosome core particle: evidence for regions of altered hydrogen bonding. *Biochemistry*, **24**, 7037-7044.
84. Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) Structure of the nucleosome core particle at 7 Å resolution. *Nature*, **311**, 532-537.
85. Akutsu, H., Nishimoto, S. and Kyogoku, Y. (1994) Dynamic structures of intact chicken erythrocyte chromatin as studied by 1H-31P cross-polarization NMR. *Biophys J*, **67**, 804-811.

86. Nishimoto, S., Akutsu, H. and Kyogoku, Y. (1987) The presence of the 30 nm filament structure of chromatin in intact chicken erythrocytes observed by ³¹P NMR. *FEBS Lett*, **213**, 293-296.
87. Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K. and Zarbock, J. (1987) The polypeptide fold of the globular domain of histone H5 in solution. A study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J*, **6**, 1833-1842.
88. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. and Sweet, R.M. (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**, 219-223.
89. Cerf, C., Lippens, G., Ramakrishnan, V., Muyldermans, S., Segers, A., Wyns, L., Wodak, S.J. and Hallenga, K. (1994) Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: full assignment, tertiary structure, and comparison with the globular domain of histone H5. *Biochemistry*, **33**, 11079-11086.
90. Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K. and Zhou, M.M. (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, **399**, 491-496.
91. Jacobs, S.A. and Khorasanizadeh, S. (2002) Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science*, **295**, 2080-2083.
92. Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V. and Laue, E.D. (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature*, **416**, 103-107.
93. Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W. and Richmond, T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol*, **319**, 1097-1113.
94. Schalch, T., Duda, S., Sargent, D.F. and Richmond, T.J. (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, **436**, 138-141.
95. Barbera, A.J., Chodaparambil, J.V., Kelley-Clarke, B., Joukov, V., Walter, J.C., Luger, K. and Kaye, K.M. (2006) The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science*, **311**, 856-861.
96. Makde, R.D., England, J.R., Yennawar, H.P. and Tan, S. (2010) Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature*, **467**, 562-566.
97. Kato, H., van Ingen, H., Zhou, B.-R., Feng, H., Bustin, M., Kay, L.E. and Bai, Y. (2011) Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proceedings of the National Academy of Sciences*, **108**, 12283.
98. Gao, M., Nadaud, P.S., Bernier, M.W., North, J.A., Hammel, P.C., Poirier, M.G. and Jaroniec, C.P. (2013) Histone H3 and H4 N-Terminal Tails in Nucleosome Arrays at Cellular Concentrations Probed by Magic Angle Spinning NMR Spectroscopy. *Journal of the American Chemical Society*, **135**, 15278-15281.
99. Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P. and Li, G. (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science*, **344**, 376-380.
100. Moriwaki, Y., Yamane, T., Ohtomo, H., Ikeguchi, M., Kurita, J.-i., Sato, M., Nagadoi, A., Shimojo, H. and Nishimura, Y. (2016) Solution structure of the isolated histone H2A-H2B heterodimer. *Scientific Reports*, **6**, 24999.

101. Shi, X., Prasanna, C., Nagashima, T., Yamazaki, T., Pervushin, K. and Nordenskiöld, L. (2018) Structure and Dynamics in the Nucleosome Revealed by Solid-State NMR. *Angew Chem Int Ed Engl*, **57**, 9734-9738.
102. Xiang, S., le Paige, U.B., Horn, V., Houben, K., Baldus, M. and van Ingen, H. (2018) Site-Specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl*, **57**, 4571-4575.
103. Bradbury, E.M., Carpenter, B.G. and Rattle, H.W.E. (1973) Magnetic Resonance Studies of Deoxyribonucleoprotein. *Nature*, **241**, 123.
104. Bradbury, E.M., Cary, P.D., Chapman, G.E., Crane-Robinson, C., Danby, S.E., Rattle, H.W., Boublík, M., Palau, J. and Aviles, F.J. (1975) Studies on the role and mode of operation of the very-lysine-rich histone H1 (F1) in eukaryote chromatin. The conformation of histone H1. *Eur J Biochem*, **52**, 605-613.
105. Bradbury, E.M., Chapman, G.E., Danby, S.E., Hartman, P.G. and Riches, P.L. (1975) Studies on the role and mode of operation of the very-lysine-rich histone H1 (F1) in eukaryote chromatin. The properties of the N-terminal and C-terminal halves of histone H1. *Eur J Biochem*, **57**, 521-528.
106. Bradbury, E.M., Danby, S.E., Rattle, H.W. and Giancotti, V. (1975) Studies on the role and mode of operation of the very-lysine-rich histone H1 (F1) in eukaryote chromatin. Histone H1 in chromatin and in H1 - DNA complexes. *Eur J Biochem*, **57**, 97-105.
107. Chapman, G.E., Hartman, P.G. and Bradbury, E.M. (1976) Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The isolation of the globular and non-globular regions of the histone H1 molecule. *Eur J Biochem*, **61**, 69-75.
108. Hartman, P.G., Chapman, G.E., Moss, T. and Bradbury, E.M. (1977) Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The three structural regions of the histone H1 molecule. *Eur J Biochem*, **77**, 45-51.
109. Aviles, F.J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C. and Bradbury, E.M. (1978) The conformation of histone H5. Isolation and characterisation of the globular segment. *Eur J Biochem*, **88**, 363-371.
110. Tancredi, T. and Temussi, P.A. (1979) Secondary structure of calf-thymus histone H1 by means of ¹³C-NMR spectroscopy. *Biopolymers*, **18**, 1-7.
111. Aviles, F.J., Danby, S.E., Chapman, G.E., Crane-Robinson, C. and Bradbury, E.M. (1979) The conformation of histone H5 bound to DNA. Maintenance of the globular structure after binding. *Biochim Biophys Acta*, **578**, 290-296.
112. Baldwin, J.P., Boseley, P.G., Bradbury, E.M. and Ibel, K. (1975) The subunit structure of the eukaryotic chromosome. *Nature*, **253**, 245-249.
113. Chapman, G.E., Aviles, F.J., Crane-Robinson, C. and Bradbury, E.M. (1978) A nuclear-magnetic-resonance study of the globular structure of the H5 histone. *Eur J Biochem*, **90**, 287-296.
114. Zarbock, J., Clore, G.M. and Gronenborn, A.M. (1986) Nuclear magnetic resonance study of the globular domain of chicken histone H5: resonance assignment and secondary structure. *Proc Natl Acad Sci U S A*, **83**, 7628-7632.
115. Wagner, G. and Wüthrich, K. (1982) Sequential resonance assignments in protein ¹H nuclear magnetic resonance spectra: Basic pancreatic trypsin inhibitor. *Journal of Molecular Biology*, **155**, 347-366.

116. Wüthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *Journal of Molecular Biology*, **155**, 311-319.
117. Cerf, C., Lippens, G., Muyldermans, S., Segers, A., Ramakrishnan, V., Wodak, S.J., Hallenga, K. and Wyns, L. (1993) Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: sequential assignment and secondary structure. *Biochemistry*, **32**, 11345-11351.
118. Garrard, W.T., Pearson, W.R., Wake, S.K. and Bonner, J. (1974) Stoichiometry of chromatin proteins. *Biochemical and Biophysical Research Communications*, **58**, 50-57.
119. Goodwin, G.H., Sanders, C. and Johns, E.W. (1973) A New Group of Chromatin-Associated Proteins with a High Content of Acidic and Basic Amino Acids. *European Journal of Biochemistry*, **38**, 14-19.
120. Abercrombie, B.D., Kneale, G.G., Crane-Robinson, C., Bradbury, E.M., Goodwin, G.H., Walker, J.M. and Johns, E.W. (1978) Studies on the conformational properties of the high-mobility-group chromosomal protein HMG 17 and its interaction with DNA. *Eur J Biochem*, **84**, 173-177.
121. Cary, P.D., King, D.S., Crane-Robinson, C., Bradbury, E.M., Rabbani, A., Goodwin, G.H. and Johns, E.W. (1980) Structural studies on two high-mobility-group proteins from calf thymus, HMG-14 and HMG-20 (ubiquitin), and their interaction with DNA. *Eur J Biochem*, **112**, 577-580.
122. Cook, G.R., Minch, M., Schroth, G.P. and Bradbury, E.M. (1989) Analysis of the binding of high mobility group protein 17 to the nucleosome core particle by ¹H NMR spectroscopy. *J Biol Chem*, **264**, 1799-1803.
123. Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M. and Luger, K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol*, **375**, 23-44.
124. Klinker, H., Haas, C., Harrer, N., Becker, P.B. and Mueller-Planitz, F. (2014) Rapid Purification of Recombinant Histones. *PLOS ONE*, **9**, e104029.
125. Hu, Q., Botuyan, M.V., Cui, G., Zhao, D. and Mer, G. (2017) Mechanisms of Ubiquitin-Nucleosome Recognition and Regulation of 53BP1 Chromatin Recruitment by RNF168/169 and RAD18. *Mol Cell*, **66**, 473-487.
126. Zhou, Z., Feng, H., Hansen, D.F., Kato, H., Luk, E., Freedberg, D.I., Kay, L.E., Wu, C. and Bai, Y. (2008) NMR structure of chaperone Chz1 complexed with histones H2A.Z-H2B. *Nature Structural & Molecular Biology*, **15**, 868.
127. Tugarinov, V. and Kay, L.E. (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR*, **28**, 165-172.
128. Kerfah, R., Plevin, M.J., Sounier, R., Gans, P. and Boisbouvier, J. (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Current Opinion in Structural Biology*, **32**, 113-122.
129. Howard, C.J., Yu, R.R., Gardner, M.L., Shimko, J.C. and Ottesen, J.J. (2015) In Liu, L. (ed.), *Protein Ligation and Total Synthesis II*. Springer International Publishing, Cham, pp. 193-226.
130. van Nuland, R., van Schaik, F.M.A., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.T.M. and van Ingen, H. (2013) Nucleosomal DNA binding

- drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics & Chromatin*, **6**, 12.
131. Xu, C., Cui, G., Botuyan, M.V. and Mer, G. (2008) Structural Basis for the Recognition of Methylated Histone H3K36 by the Eaf3 Subunit of Histone Deacetylase Complex Rpd3S. *Structure*, **16**, 1740-1750.
 132. Munari, F., Soeroes, S., Zenn, H.M., Schomburg, A., Kost, N., Schröder, S., Klingberg, R., Rezaei-Ghaleh, N., Stützer, A., Gelato, K.A. *et al.* (2012) Methylation of Lysine 9 in Histone H3 Directs Alternative Modes of Highly Dynamic Interaction of Heterochromatin Protein hHP1 β with the Nucleosome. *Journal of Biological Chemistry*, **287**, 33756-33765.
 133. Kitevski-LeBlanc, J., Fradet-Turcotte, A., Kukic, P., Wilson, M.D., Portella, G., Yuwen, T., Panier, S., Duan, S., Canny, M.D., van Ingen, H. *et al.* (2017) The RNF168 paralog RNF169 defines a new class of ubiquitylated histone reader involved in the response to DNA damage. *eLife*, **6**, e23872.
 134. Dorigo, B., Schalch, T., Bystricky, K. and Richmond, T.J. (2003) Chromatin Fiber Folding: Requirement for the Histone H4 N-terminal Tail. *Journal of Molecular Biology*, **327**, 85-96.
 135. Wiesner, S. and Sprangers, R. (2015) Methyl groups as NMR probes for biomolecular interactions. *Curr Opin Struct Biol*, **35**, 60-67.
 136. Ollerenshaw, J.E., Tugarinov, V. and Kay, L.E. (2003) Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry*, **41**, 843-852.
 137. Keizers, P.H. and Ubbink, M. (2011) Paramagnetic tagging for protein structure and dynamics analysis. *Prog Nucl Magn Reson Spectrosc*, **58**, 88-96.
 138. Amero, C., Schanda, P., Durá, M.A., Ayala, I., Marion, D., Franzetti, B., Brutscher, B. and Boisbouvier, J. (2009) Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *Journal of the American Chemical Society*, **131**, 3448-3449.
 139. Macek, P., Kerfah, R., Boeri Erba, E., Crublet, E., Moriscot, C., Schoehn, G., Amero, C. and Boisbouvier, J. (2017) Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Science Advances*, **3**, e1601601.
 140. Sprangers, R. and Kay, L.E. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*, **445**, 618.
 141. Sapienza, P.J. and Lee, A.L. (2010) Using NMR to study fast dynamics in proteins: methods and applications. *Current Opinion in Pharmacology*, **10**, 723-730.
 142. Korzhnev, D.M., Kloiber, K., Kanelis, V., Tugarinov, V. and Kay, L.E. (2004) Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. *Journal of the American Chemical Society*, **126**, 3964-3973.
 143. Yuwen, T., Huang, R. and Kay, L.E. (2017) Probing slow timescale dynamics in proteins using methyl ¹H CEST. *Journal of Biomolecular NMR*, **68**, 215-224.
 144. Rennella, E., Huang, R., Velyvis, A. and Kay, L.E. (2015) ¹³CHD2-CEST NMR spectroscopy provides an avenue for studies of conformational exchange in high molecular weight proteins. *Journal of Biomolecular NMR*, **63**, 187-199.

145. Bertini, I., Luchinat, C., Parigi, G., Ravera, E., Reif, B. and Turano, P. (2011) Solid-state NMR of proteins sedimented by ultracentrifugation. *Proceedings of the National Academy of Sciences*, **108**, 10396.
146. Gardiennet, C., Schütz, A.K., Hunkeler, A., Kunert, B., Terradot, L., Böckmann, A. and Meier, B.H. (2012) A Sedimented Sample of a 59 kDa Dodecameric Helicase Yields High-Resolution Solid-State NMR Spectra. *Angewandte Chemie International Edition*, **51**, 7855-7858.
147. Munowitz, M.G. and Griffin, R.G. (1982) Two-dimensional nuclear magnetic resonance in rotating solids: An analysis of line shapes in chemical shift-dipolar spectra. *The Journal of Chemical Physics*, **76**, 2848-2858.
148. van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J. and Bonvin, A.M.J.J. (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, **428**, 720-725.
149. Karaca, E., Rodrigues, J.P.G.L.M., Graziadei, A., Bonvin, A.M.J.J. and Carlomagno, T. (2017) M3: an integrative framework for structure determination of molecular machines. *Nature Methods*, **14**, 897.
150. Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B. and Sali, A. (2012) Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLOS Biology*, **10**, e1001244.
151. van Ingen, H. and Bonvin, A.M.J.J. (2014) Information-driven modeling of large macromolecular assemblies using NMR data. *Journal of Magnetic Resonance*, **241**, 103-114.
152. Tamò, G.E., Abriata, L.A. and Dal Peraro, M. (2015) The importance of dynamics in integrative modeling of supramolecular assemblies. *Current Opinion in Structural Biology*, **31**, 28-34.
153. Horn, V. and van Ingen, H. (2018) In Logie, C. (ed.), *Epigenetics & Chromatin*, InTechOpen.
154. Xu, X., Ben Imeddourene, A., Zargarian, L., Foloppe, N., Mauffret, O. and Hartmann, B. (2014) NMR Studies of DNA Support the Role of Pre-Existing Minor Groove Variations in Nucleosome Indirect Readout. *Biochemistry*, **53**, 5601-5612.
155. Imeddourene, A.B., Xu, X., Zargarian, L., Oguey, C., Foloppe, N., Mauffret, O. and Hartmann, B. (2016) The intrinsic mechanics of B-DNA in solution characterized by NMR. *Nucleic Acids Research*, **44**, 3432-3447.
156. Hammond, C.M., Strømme, C.B., Huang, H., Patel, D.J. and Groth, A. (2017) Histone chaperone networks shaping chromatin function. *Nature Reviews Molecular Cell Biology*, **18**, 141.
157. Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J.M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K.K., Lemak, A. *et al.* (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences*, **105**, 4685.
158. Mousson, F., Lautrette, A., Thuret, J.-Y., Agez, M., Courbeyrette, R., Amigues, B., Becker, E., Neumann, J.-M., Guerois, R., Mann, C. *et al.* (2005) Structural basis for the interaction of Asf1 with histone H3 and its functional implications. *Proceedings of the National Academy of Sciences*, **102**, 5975.

159. Su, D., Hu, Q., Zhou, H., Thompson, J.R., Xu, R.-M., Zhang, Z. and Mer, G. (2011) Structure and Histone Binding Properties of the Vps75-Rtt109 Chaperone-Lysine Acetyltransferase Complex. *Journal of Biological Chemistry*, **286**, 15625-15629.
160. Hansen, D.F., Zhou, Z., Feng, H., Jenkins, L.M.M., Bai, Y. and Kay, L.E. (2009) Binding Kinetics of Histone Chaperone Chz1 and Variant Histone H2A.Z-H2B by Relaxation Dispersion NMR Spectroscopy. *Journal of Molecular Biology*, **387**, 1-9.
161. Zhou, Z., Feng, H., Zhou, B.R., Ghirlando, R., Hu, K., Zwolak, A., Miller Jenkins, L.M., Xiao, H., Tjandra, N., Wu, C. *et al.* (2011) Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature*, **472**, 234-237.
162. Warren, C., Matsui, T., Karp, J.M., Onikubo, T., Cahill, S., Brenowitz, M., Cowburn, D., Girvin, M. and Shechter, D. (2017) Dynamic intramolecular regulation of the histone chaperone nucleoplasmin controls histone binding and release. *Nat Commun*, **8**, 2215.
163. Corbeski, I., Dolinar, K., Wienk, H., Boelens, R. and van Ingen, H. (2018) DNA repair factor APLF acts as a H2A-H2B histone chaperone through binding its DNA interaction surface. *Nucleic Acids Res*, **46**, 7138-7152.
164. Borgia, A., Borgia, M.B., Bugge, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D. *et al.* (2018) Extreme disorder in an ultrahigh-affinity protein complex. *Nature*, **555**, 61-66.
165. Zhou, B.R., Feng, H.Q., Kato, H., Dai, L., Yang, Y.D., Zhou, Y.Q. and Bai, Y.W. (2013) Structural insights into the histone H1-nucleosome complex. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 19390-19395.
166. Kitevski-LeBlanc, J.L., Yuwen, T., Dyer, P.N., Rudolph, J., Luger, K. and Kay, L.E. (2018) Investigating the Dynamics of Destabilized Nucleosomes Using Methyl-TROSY NMR. *J Am Chem Soc*, **140**, 4774-4777.
167. Zhou, B.R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T.S. and Bai, Y. (2015) Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell*, **59**, 628-638.
168. Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A.R., Papai, G., Reymer, A., Syed, S.H., Lone, I.N., Tonchev, O., Crucifix, C. *et al.* (2017) Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Molecular Cell*, **66**, 384-397.
169. Zhou, B.-R., Feng, H., Ghirlando, R., Li, S., Schwieters, C.D. and Bai, Y. (2016) A Small Number of Residues Can Determine if Linker Histones Are Bound On or Off Dyad in the Chromatosome. *Journal of Molecular Biology*, **428**, 3948-3959.
170. Turner, A.L., Watson, M., Wilkins, O.G., Cato, L., Travers, A., Thomas, J.O. and Stott, K. (2018) Highly disordered histone H1-DNA model complexes and their condensates. *Proc Natl Acad Sci U S A*, **115**, 11964-11969.
171. Zhou, B.-R., Feng, H., Ghirlando, R., Kato, H., Gruschus, J. and Bai, Y. (2012) Histone H4 K16Q Mutation, an Acetylation Mimic, Causes Structural Disorder of Its N-Terminal Basic Patch in the Nucleosome. *Journal of Molecular Biology*, **421**, 30-37.
172. Kato, H., Gruschus, J., Ghirlando, R., Tjandra, N. and Bai, Y. (2009) Characterization of the N-Terminal Tail Domain of Histone H3 in Condensed Nucleosome Arrays by Hydrogen Exchange and NMR. *Journal of the American Chemical Society*, **131**, 15104-15105.
173. Stützer, A., Liokatis, S., Kiesel, A., Schwarzer, D., Sprangers, R., Söding, J., Selenko, P. and Fischle, W. (2016) Modulations of DNA Contacts by Linker Histones and Post-

- translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Molecular Cell*, **61**, 247-259.
174. Debelouchina, G.T., Gerecht, K. and Muir, T.W. (2017) Ubiquitin utilizes an acidic surface patch to alter chromatin structure. *Nat Chem Biol*, **13**, 105-110.
 175. Monneau, Y.R., Soufari, H., Nelson, C.J. and Mackereth, C.D. (2013) Structure and activity of the peptidyl-prolyl isomerase domain from the histone chaperone Fpr4 toward histone H3 proline isomerization. *J Biol Chem*, **288**, 25826-25837.
 176. Dose, A., Liokatis, S., Theillet, F.X., Selenko, P. and Schwarzer, D. (2011) NMR profiling of histone deacetylase and acetyl-transferase activities in real time. *ACS Chem Biol*, **6**, 419-424.
 177. Liokatis, S., Klingberg, R., Tan, S. and Schwarzer, D. (2016) Differentially Isotope-Labeled Nucleosomes To Study Asymmetric Histone Modification Crosstalk by Time-Resolved NMR Spectroscopy. *Angewandte Chemie International Edition*, **55**, 8262-8265.
 178. Sinha, K.K., Gross, J.D. and Narlikar, G.J. (2017) Distortion of histone octamer core promotes nucleosome mobilization by a chromatin remodeler. *Science*, **355**, eaaa3761.
 179. Speranzini, V., Pilotto, S., Sixma, T.K. and Mattevi, A. (2016) Touch, act and go: landing and operating on nucleosomes. *EMBO J*, **35**, 376-388.
 180. Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eisenberg, J.C., Allis, C.D. and Khorasanizadeh, S. (2001) Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J*, **20**, 5232-5241.
 181. Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41-45.
 182. Turner, B.M. (2000) Histone acetylation and an epigenetic code. *Bioessays*, **22**, 836-845.
 183. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074-1080.
 184. Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H. and Allis, C.D. (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*, **438**, 1116-1122.
 185. Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J. and Zhou, M.M. (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature*, **466**, 258-262.
 186. Nady, N., Lemak, A., Walker, J.R., Avvakumov, G.V., Kareta, M.S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X. *et al.* (2011) Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J Biol Chem*, **286**, 24300-24311.
 187. Kostrohn, S., Kontaxis, G., Kaufmann, T., Schirghuber, E., Kubicek, S., Konrat, R. and Slade, D. (2017) A histone-mimicking interdomain linker in a multidomain protein modulates multivalent histone binding. *J Biol Chem*, **292**, 17643-17657.
 188. Perell, G.T., Mishra, N.K., Sudhamalla, B., Ycas, P.D., Islam, K. and Pomerantz, W.C.K. (2017) Specific Acetylation Patterns of H2A.Z Form Transient Interactions with the BPTF Bromodomain. *Biochemistry*, **56**, 4607-4615.

189. Cato, L., Stott, K., Watson, M. and Thomas, J.O. (2008) The interaction of HMGB1 and linker histones occurs through their acidic and basic tails. *J Mol Biol*, **384**, 1262-1272.
190. Gatchalian, J., Wang, X., Ikebe, J., Cox, K.L., Tencer, A.H., Zhang, Y., Burge, N.L., Di, L., Gibson, M.D., Musselman, C.A. *et al.* (2017) Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nat Commun*, **8**, 1489.
191. Morrison, E.A., Bowerman, S., Sylvers, K.L., Wereszczynski, J. and Musselman, C.A. (2018) The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife*, **7**, e31481.
192. Musselman, C.A., Gibson, M.D., Hartwick, E.W., North, J.A., Gatchalian, J., Poirier, M.G. and Kutateladze, T.G. (2013) Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun*, **4**, 2969.
193. Richart, A.N., Brunner, C.I.W., Stott, K., Murzina, N.V. and Thomas, J.O. (2012) Characterization of Chromoshadow Domain-mediated Binding of Heterochromatin Protein 1 α (HP1 α) to Histone H3. *Journal of Biological Chemistry*, **287**, 18730-18737.
194. Fenley, A.T., Anandakrishnan, R., Kidane, Y.H. and Onufriev, A.V. (2018) Modulation of nucleosomal DNA accessibility via charge-altering post-translational modifications in histone core. *Epigenetics Chromatin*, **11**, 11.
195. Bowman, G.D. and Poirier, M.G. (2015) Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chemical Reviews*, **115**, 2274-2295.
196. Tessarz, P. and Kouzarides, T. (2014) Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews Molecular Cell Biology*, **15**, 703-708.
197. Kato, H., Jiang, J., Zhou, B.R., Rozendaal, M., Feng, H., Ghirlando, R., Xiao, T.S., Straight, A.F. and Bai, Y. (2013) A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science*, **340**, 1110-1113.
198. Nelissen, F.H.T., Tessari, M., Wijmenga, S.S. and Heus, H.A. (2016) Stable isotope labeling methods for DNA. *Prog Nucl Magn Reson Spectrosc*, **96**, 89-108.
199. Shepherd, N.E., Gamsjaeger, R., Vandevenne, M., Cubeddu, L. and Mackay, J.P. (2015) Site directed nitroxide spin labeling of oligonucleotides for NMR and EPR studies. *Tetrahedron*, **71**, 813-819.
200. Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M. *et al.* (2018) The interaction landscape between transcription factors and the nucleosome. *Nature*, **562**, 76-81.

2

ARCHITECTURAL PRINCIPLES OF
ARCHAEAL AND EUKARYOTIC CHROMATIN
COMPLEXES

Based on the review article: B. Henneman, **C. van Emmerik**, H. van Ingen and R.T. Dame (2018) Structure and function of archaeal histones. *PLoS Genet*, **14**, e1007582

Based on the research article: B. Henneman*, T. Brouwer*, **C. van Emmerik**, R.A. van der Valk, M. Timmer, N. Kirolos, H. van Ingen, J. van Noort and R.T. Dame (2019) Mechanical and structural properties of archaeal hypernucleosomes. *Submitted*.

*These authors contributed equally.

Multiple sequence alignment of new archaeal histones and Tethered Particle Motion (TPM) experiments were performed by Bram Henneman at the Leiden Institute for Chemistry, Leiden University.

Magnetic Tweezer force experiments were performed by Thomas Brouwer at the Department of Biological and Soft Matter Physics, Huygens-Kamerlingh Onnes Laboratory, Leiden University.

ABSTRACT

Histones are primarily responsible for the compaction of genomic DNA into chromatin, both in eukaryotes and in most archaea. As eukaryotes are classified as being part of the archaeal evolutionary branch, chromatin organization as we have come to understand it in eukaryotes may have evolved from that of the archaeal lineage. Recently, the archaeal histones HMfA and HMfB have been shown to form a continuous histone-DNA complex, termed the hypernucleosome. Since the vast majority of newly identified archaeal histones have only been identified at the sequence level, we here used computational modeling to analyze their propensity to form hypernucleosomes. Sequence analysis and structural modeling suggest that the shape, symmetry and surface residues of the histones dictate the type of nucleosome-like complex that can be formed with genomic DNA, as well as its stability and level of compaction. In addition, modeling of archaeal histone complexes highlighted the fundamental design principles of the eukaryotic nucleosome, demonstrating why eukaryotic histones are not able to form a hypernucleosome. We propose that different archaeal histone dimers in the same organism can specifically promote or disturb hypernucleosome formation and that this might be functional in gene regulation. Differentiation of histones into asymmetric dimers and the emergence of modifiable N-terminal tails may have led to the evolution of eukaryotic octameric nucleosomes and an epigenetic code providing extensive regulation of DNA-templated processes.

INTRODUCTION

Architectural chromatin proteins are found in every domain of life. In eukaryotes and most archaeal lineages, histones are primarily responsible for the compaction of genomic DNA into chromatin. Comparative genome studies have demonstrated that bacteria and archaea share a common ancestor; eukaryotes are to date classified as being part of the archaeal branch (1-3). This raises the intriguing possibility that chromatin organization as we have come to understand it in eukaryotes has evolved from that of the archaeal lineage. In this chapter we will explore this idea by studying the architectural principles of archaeal and eukaryotic chromatin complexes on a structural level using computational modeling.

The eukaryotic nucleosome is composed of ~147 bp of DNA wrapped around an octameric histone core (4). Histone proteins display the characteristic histone fold, featuring a long α -helix enclosed by two shorter ones, and associate to form heterodimers (H2A/H2B) and -tetramers ((H3/H4)₂) before assembling into octamers. Histone dimers associate by forming a 4-helix bundle (4HB) featuring a highly conserved salt bridge and several hydrophobic interactions. Each histone dimer has three contact points with the surrounding DNA, wrapping it around the histone core in ~1.6 superhelical turns. In almost all cases these histone-DNA contacts involve an arginine residue protruding into the inward facing DNA minor groove (Fig. 1). The dimer-dimer contacts and the dimer-DNA contacts are essential for holding the nucleosome complex together. Apart from the folded domain, histones contain N- and/or C-terminal tails that protrude from the octameric core, partly through the DNA gyres. These tails are an important platform for the deposition of post-translational modifications that are part of the epigenetic code (5). The tails are very basic and can therefore interact specifically with close-by DNA on the same or a neighboring nucleosome. In addition, the N-terminal tail of histone H4 is known to be able to bind the acidic patch on the surface of an adjacent nucleosome, facilitating compaction (4,6). Hence, the histone tails are at several levels crucial for regulation of chromatin function.

Archaeal genomes encode several types of DNA-binding proteins that are involved in compaction of genomic DNA, including bacterial HU homologues, Alba family proteins

as well as histones, which are found in many archaeal species throughout the domain (7,8). Characteristic of these proteins is their ability to bend and bridge DNA as well as to form DNA filaments. The most well-studied archaeal histones are HMfA and HMfB, found in *Methanothermobacter fervidus* (9). Sequence identity with eukaryotic histones is limited, but NMR and X-ray studies showed that the 3D structure of these archaeal histones is identical to eukaryotic histones, although they are lacking N- and C-terminal tails (Fig. 1) (10,11). HMf dimers can be homodimeric or heterodimeric and the two histone variants are expressed at different levels and ratios in different growth phases, suggesting a distinct function for both proteins (12). Several studies have reported on how archaeal histones could bind and compact DNA, as they do not seem to form octameric nucleosome-like complexes. Although some papers conclude that the dimers bind separately or as tetramers (13,14), later observations based on MNase assays of *Thermococcus kodakarensis* chromatin led to the idea that archaeal histone dimers multimerize and wrap DNA into a filament of variable length (15). This compaction mechanism was confirmed by the crystal structure determined by Luger and coworkers at 4 Å resolution (16), showing that HMfB dimers assemble into an endless left-handed rod *in vitro*, the so-called hypernucleosome.

The vast majority of archaeal histones have only been identified at the sequence level and their structures are unknown (17). In addition, new archaeal species encoding histones are still being discovered. The lack of tails in most archaeal histones and clear difference in formation of histone-DNA complexes shown for HMfB suggests fundamentally different mechanisms of compaction and regulation of DNA-templated processes in archaea. In order to obtain a better understanding of archaeal chromatin, we aimed to build structural models of archaeal histones and histone complexes to compare them to each other and to eukaryotic systems. At the same time, this yielded fundamental insights into the architecture of the eukaryotic nucleosome, demonstrating why eukaryotic histones are not able to form a hypernucleosome. We find that the shape, symmetry and interacting surface residues of the histone dimers dictate the type of nucleosome-like complex that can be formed with genomic DNA.

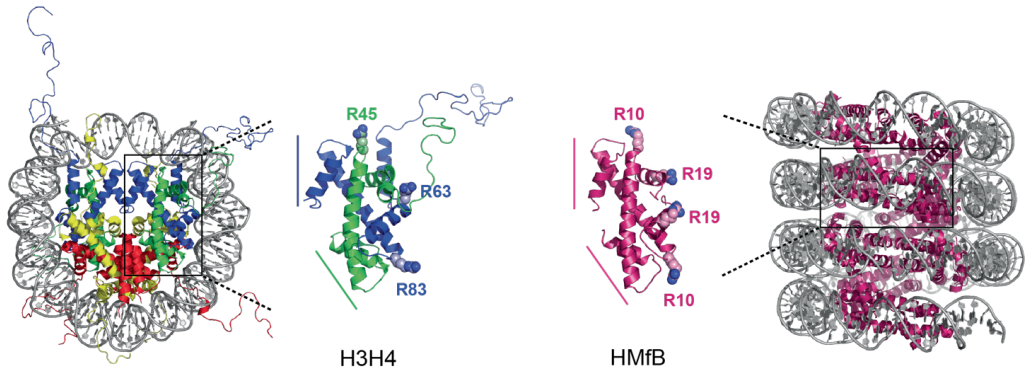


Figure 1. Eukaryotic and archaeal histone complexes. The eukaryotic nucleosome (PDB entry: 1KX5) consists of two copies of H2A, H2B, H3 and H4 forming an octameric core, around which ~147 bp of DNA is wrapped (left panel). Histones H3 and H4 form heterodimers with two distinct dimerization interfaces (indicated by colored lines), one to interact with another H3H4-dimer (blue) and another to interact with H2AH2B (green). Eukaryotic histones contain long N-terminal tails. The archaeal hypernucleosome (PDB entry: 5T5K) is composed of HMfB homodimers that can assemble into a multimer of variable length (right panel) wrapping the DNA into a continuous superhelix. The HMfB dimers are symmetric and therefore feature two identical dimerization interfaces (pink lines). These archaeal histones do not have N-terminal tails. For both dimers, DNA-anchoring arginine residues are shown as spheres and labelled. In both cases three DNA contact points are formed by these arginines.

RESULTS

We used a multi-step modeling approach to construct a model of the HMfB hypernucleosome, building the HMfB dimer and the hypernucleosomal DNA separately and positioning the HMfB dimers onto the DNA using the eukaryotic nucleosome structure as a template. We aimed to make the model as symmetric and regular as possible, as this seems to be the most likely way a multimer can be formed on DNA if only one type of histone dimer is involved. During the course of this work, the crystal structure of the hypernucleosome solved by the Luger lab became available and was thus used to evaluate the model.

MODEL BUILDING

The first step in building an HMfB hypernucleosome was to obtain a symmetric dimer of HMfB to be used as a repeating unit in the histone multimer core. Dimer symmetry

will ensure identical dimer-dimer interfaces throughout the multimer and therefore regularity of the complex. The crystallized HMfB dimer structure is fully symmetric (10), however, formation of a nucleosome-like multimer is not possible with this dimer because the $\alpha 3$ -helices come too close together in a tetramer conformation. Arginine 66 in the $\alpha 3$ -helix causes a clash in the 4HB (Fig. 2A) that cannot be resolved by repositioning the side chains only.

To position the C-terminal residues of HMfB in a way that prevents the arginines from clashing, the eukaryotic histone H3 structure inside the nucleosome was used as a template for remodeling of HMfB. H3 and HMfB have an overall backbone RMSD of 3.2 Å, but the structural differences are found mainly in the N- and C-terminal residues of the $\alpha 1$ - and $\alpha 3$ -helices, respectively (backbone RMSD over residues 18-60 of HMfB is 0.5 Å). As histone H3 forms a stable 4HB interface with the other H3 copy in the nucleosome, modeling the HMfB $\alpha 3$ -helix according to this structure was expected to yield an HMfB molecule that can do the same. Notably, modeling HMfB according to H3 *and* H4 as a dimer may yield an even more similar 4HB interface, but it would not be likely to form a regular multimer due to its asymmetry (Fig. 1).

To construct the symmetric H3-templated HMfB dimer, two copies of the modelled HMfB monomer were superimposed on the HMfB dimer crystal structure to maintain the native dimer orientation. Comparison of the resulting HMfB dimer to the crystal structure shows that the last turn of the $\alpha 3$ -helix is unwound in order to make space for arginine R66 in the 4HB (Fig. 2A). Positioning of the modelled HMfB dimers in tetramer conformation shows that the arginine clash is indeed resolved, and a stable interface can be formed.

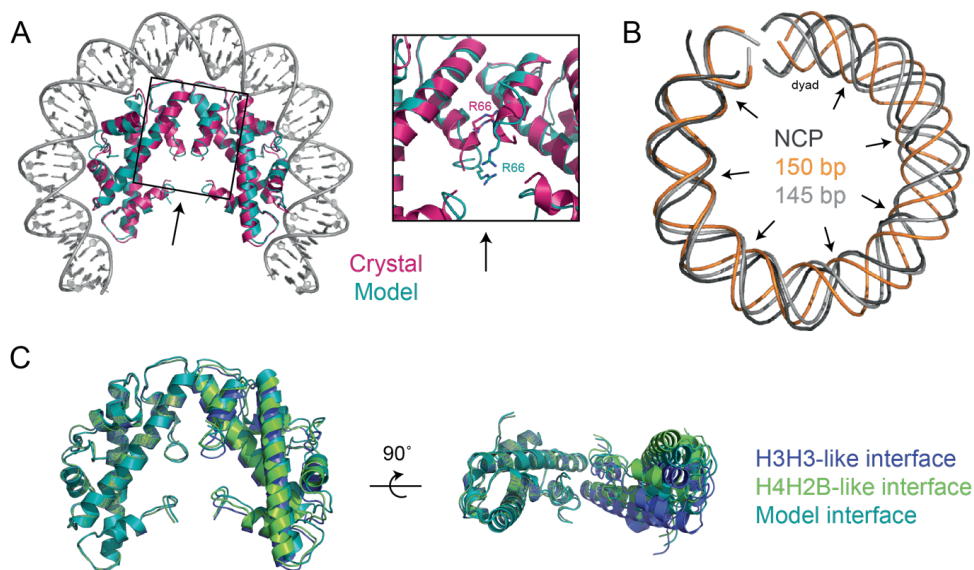


Figure 2. Building an HMfB hypernucleosome model. (A) Two HMfB dimer structures as determined by X-ray (PDB entry: 1A7W) in H3H4-like tetramer conformation lead to an arginine clash in the 4HB (zoomed view, R66 indicated). H3-templated remodeling of HMfB resolves this clash by unwinding the C-terminal turn of the α 3-helix. HMfB dimer crystal structure in pink, modelled HMfB dimer in teal. (B) Stretching of the modelled DNA superhelix by shortening the sequence from 150 bp to 145 bp while keeping the total bend angle the same enables alignment of the DNA minor grooves with the nucleosomal DNA. Nucleosomal DNA (PDB entry: 1KX5) in black, 150 bp DNA model in orange, 145 bp DNA model (stretched) in grey. Minor grooves are indicated with arrows. (C) Comparison of nucleosomal interfaces: HMfB dimer models were placed according to either the H3H3 4HB (blue) or the H4H2B 4HB (green) in the nucleosome and compared to the dimer orientation in the final hypernucleosome model (teal).

In the next step, we constructed a hypernucleosomal DNA molecule with defined superhelical parameters to fit the HMfB dimers. The eukaryotic nucleosomal DNA is \sim 147 bp long and has very irregular bend and twist values, featuring stretched regions and kinking (18), which makes it hard to extend it for hypernucleosome construction. Therefore, an ideal superhelix was built using optimized base-pair step parameters to match the diameter, superhelical pitch and register of the nucleosomal DNA as closely as possible.

The eukaryotic nucleosomal DNA is generally B-form and has an average bend angle of 4.53° per base pair (19), which determines the diameter of the superhelix (\sim 80 Å).

Therefore, this angle was used to bend standard B-form DNA to obtain an initial model of an ideal nucleosomal superhelix. The next step was to adjust the vertical distance between the DNA gyres, the superhelical pitch, to match the pitch of the nucleosomal DNA. The pitch of the superhelix is inversely correlated to the twist value, a base pair step parameter describing the angle between the successive base pairs around the helical axis, which determines the number of base pairs per helical turn. For the ideal superhelix, the pitch was slightly decreased by increasing the twist value to 35.5° per bp (10.14 bp per helical turn) to match the nucleosomal pitch, thereby avoiding the histone dimers being too closely stacked in the multimer.

Interestingly, placing HMfB dimers onto the resulting superhelix showed that the inward facing minor grooves of the DNA are too close together to properly accommodate histone dimers (Fig. 2B). The distance between inward facing minor grooves has to fit the position of the anchoring arginines on the histone dimers to accomplish stable histone-DNA contacts. To accomplish this, the modeled superhelix was stretched by shortening the sequence while keeping the total bend angle over all residues constant. This increases the bend angle per base pair from 4.5° (675° over 150 bp) to 4.65° per base pair (675° over 145 bp), while maintaining the original diameter of the superhelix. This ensures that the superhelix still matches the surface curvature of the histone dimers and now also allows the arginines to anchor to the minor grooves.

Finally, we explored strategies to assemble the HMfB dimers onto the ideal superhelix to form a hypernucleosome. The first strategy was to take the nucleosomal dimer-dimer interfaces (formation of the 4HB) as a guide to construct the complex. The eukaryotic nucleosome contains two different 4HBs, one in which the two H3 histones are interacting and another in which H4 and H2B are interacting. Interestingly, these two interfaces do not have identical conformations, as the H4H2B interface has a higher pitch and a wider orientation of the dimers (Fig. 2C). As the HMfB dimer is a symmetrical multimerization unit, we assumed that each dimer-dimer interface in the multimer will be the same. Consequently, an HMfB multimer constructed according to H3H3 interface orientation is too flat to match the pitch of the DNA superhelix and a H4H2B-templated multimer is too wide. In both cases, the crucial arginine-minor groove anchoring cannot be established.

Therefore, we decided to take the nucleosomal DNA-dimer interface as a guide to position HMfB dimers onto the ideal superhelix DNA rather than the dimer-dimer interfaces. The symmetric HMfB dimer was superimposed on histone H3 of the eukaryotic nucleosome and the idealized 145 bp DNA superhelix was aligned with the nucleosomal DNA. Subsequently, the HMfB dimer including the surrounding 30 bp of ideal superhelix DNA was used as a building block to create the hypernucleosome. The 145 bp ideal superhelix was extended by copying it and aligning the two copies over 25 bp. As the modelled DNA is completely regular, they can easily be merged to yield a 265 bp DNA molecule that can accommodate eight histone dimers. The atomic coordinates of this block were copied and translated to the next HMfB dimer position by aligning the block's dyad to the next dyad on the 265 bp DNA superhelix (30 bp away). This method indeed yielded a dimer-dimer orientation that was in between the H3H3 and the H4H2B positioning (Fig. 2C). This way, eight HMfB dimers were positioned onto the ideal superhelix to form a model of the hypernucleosome. This complex was water-refined using the HADDOCK web server refinement interface (20), in order to improve the physico-chemical properties of the dimer-dimer and dimer-DNA contacts and remove side chain clashes. The resulting 4HB was indeed free of clashes and contained the expected salt bridge and hydrophobic contacts.

We next investigated whether HMfB multimerization is compatible with different diameters of the hypernucleosome. A model using a DNA superhelix with a diameter of ~ 90 Å (bend angle of 600° over 145 bp), where the nucleosome-like model has ~ 80 Å, shows fewer HMfB-DNA and HMfB-HMfB contacts. In this model, each HMfB dimer has two instead of three DNA contacts per HMfB dimer, due to insufficient curvature of the DNA to match the HMfB dimer surface. The critical salt bridge (H49-D59) in the 4HB is present only in three out of seven interfaces, because the dimers are further apart. Also, only three axial contacts per dimer are found versus eight contacts per dimer in the more compact model. In conclusion, a wider DNA superhelix is less likely to facilitate multimerization, because of the mismatch between the dimers' surface and the DNA curvature and the lack of axial dimer-dimer contacts to stabilize the complex. HMfB thus likely causes a DNA bending angle that is highly similar to the curvature in the eukaryotic nucleosome due to the intrinsic structural similarities, determining the dimensions of the nucleosome-like complex that can be formed.

The final hypernucleosome model was evaluated by comparing it to crystal structure that became available during the course of this work (16). The global features of the model and the structure are very similar (Fig. 3). The diameter, pitch and register of the DNA are identical and also the positioning of the histone dimers is highly similar. Interestingly, the α 3-helix in the 4HB interface is still intact in the crystal structure in agreement with the HMfB dimer structure, but the distance between the two helices is larger than in the H3H3-like orientation ($C\alpha$ -distances of 5.9 vs. 3.4 Å, see Fig. 2A), allowing R66 of both dimers to be accommodated without clashing. Notably, in the crystal structure, all dimers are lacking the last one or two lysines of the sequence, meaning that these might be flexible.

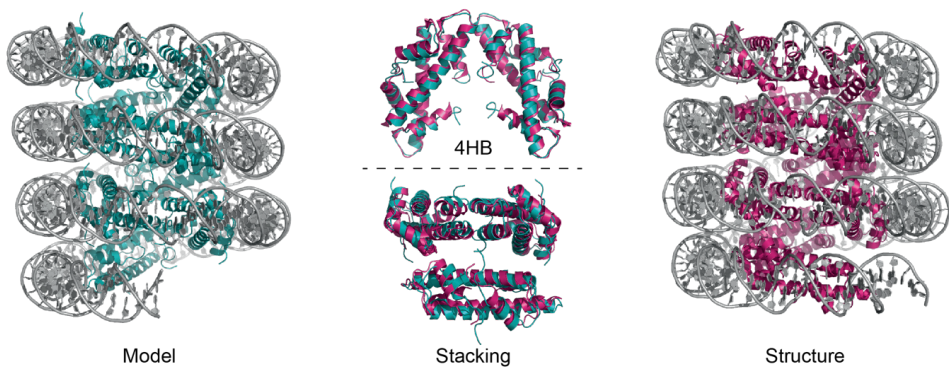


Figure 3. Comparison of the HMfB hypernucleosome model (left panel, eight HMfB dimers in teal) to the crystal structure (PDB entry: 5T5K, right panel, nine HMfB dimers in pink). In the middle panel, the 4HB interface and the stacking interface overlays are displayed.

SEQUENCE ALIGNMENT OF (NEW) ARCHAEOAL HISTONES

In recent years, metagenomic sequencing has enabled the discovery of many new lineages in the archaeal domain, including many newly identified species encoding histones in their genome. As most of these histones or histone-like proteins have not been studied *in vitro* or *in vivo* yet, molecular modeling could provide a way to select interesting sequences for future studies on structure and their potential ability to form hypernucleosomes. A selection of histone protein sequences from archaea throughout the domain was aligned together with eukaryotic histone H4 to identify similarities and differences (Fig. 4).

Dimerization of both eukaryotic histones and HMfB is driven by a hydrophobic core (involving residues A24, L28, L32, I39, and A43 in HMfB) as well as a crucial salt bridge for a stable histone fold (R52-D59 in HMfB). These hydrophobic residues and the salt bridge are conserved among archaea (Fig. 4). This indicates that archaeal histones can dimerize and form very similar tertiary structures. Also, residues that play an important role in DNA binding are present in all examined histones, including the arginines that anchor archaeal histone dimers to the DNA minor grooves (R10 and R19 in HMfB). Both eukaryotic H3H4-dimers and HMfB dimers can form tetramers by hydrogen bonding of H49 and D59 (HMfB) and additional hydrophobic interactions in the interface (L46 and L62 in HMfB), pairs of residues that, too, are generally conserved among archaeal histones. Notably, many archaeal genomes contain multiple histone sequences. This suggests that different types of homodimers and/or heterodimers can form which could be functional in transcriptional regulation. The most remarkable difference is the variety of N- and C-terminal extensions and truncations in the different histones. Interestingly, the tails of the two histones from *Heimdallarchaeota* and *Huberarchaea* are of roughly the same length and sequence composition as eukaryotic H4 tails, featuring several basic residues that in eukaryotes are important targets for post-translational modification.

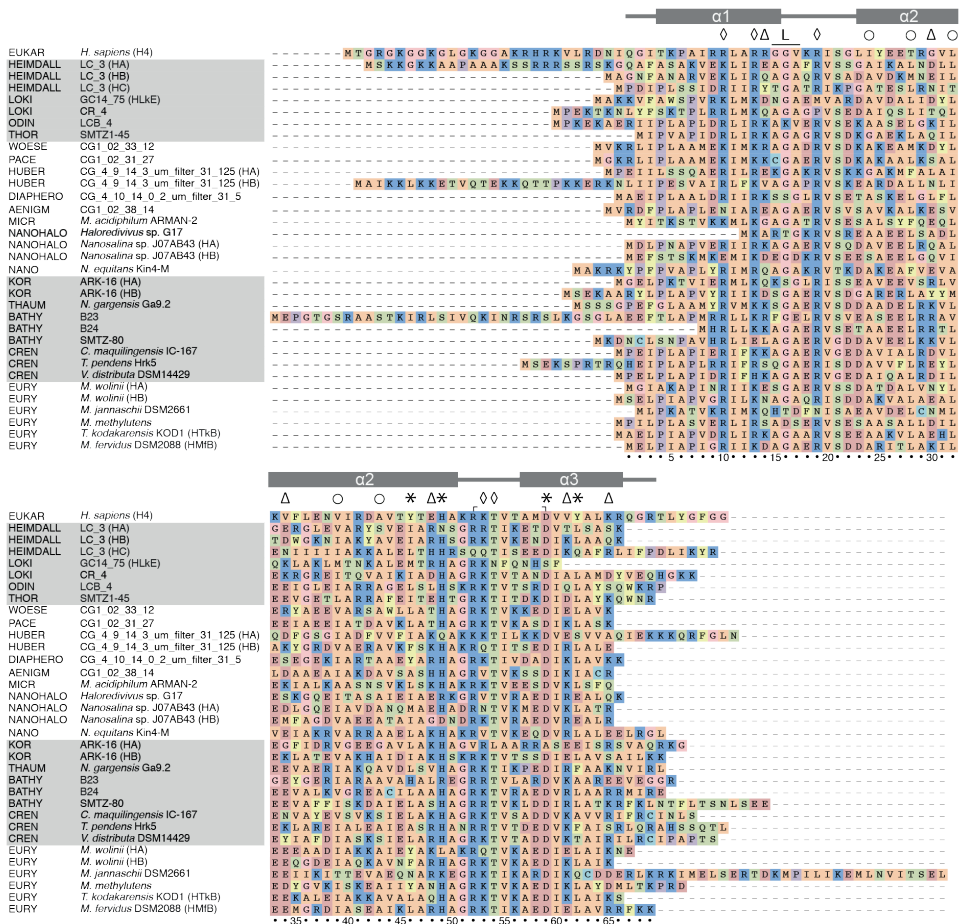


Figure 4. Sequence alignment of archaeal histones. Colors indicate the side chain group: R, H, K: blue; D, E: red; A, V, I, L, M: orange; F, Y, W: yellow; S, T, N, Q: green; C: turquoise; G: pink; P: purple. Symbols above the alignment indicate residues involved in histone dimerization (\circ and salt bridge R52–D59), histone-DNA contacts (\diamond), formation of the 4HB ($*$), the loop in the stacking interface (L), and the putative stacking interactions (Δ) based on HmFB. Secondary structure and numbering of HmFB is used for reference. Sequences are indicated by the name of the phylum, name of the species or genome and histone name (if more than one is present in one genome). They are grouped according to the four main archaeal supergroups (grey shading): Asgard archaea, DPANN archaea, TACK archaea and Euryarchaeota.

HYPERNUCLEOSOME PROPENSITY OF OTHER ARCHAEAL HISTONES

The sequence alignment was used to build straight-forward homo-dimeric homology models of the selected archaeal histones based on the structure of HMfB. These models were superimposed onto HMfB dimers in the hypernucleosome crystal structure to assess their potential to form a hypernucleosome-like structure similar to HMfB and whether alternative or additional interactions were possible in the different archaeal histone complexes.

To evaluate whether archaeal histones other than HMfB could assemble into hypernucleosomes, we defined three criteria for complex formation. Firstly, conservation of residues in the 4HB (L46, H49, D59, and L62 in HMfB) is required, as forming a tetramer is the first step in multimerization. Secondly, residue G16, which is positioned at the stacking interface of the hypernucleosome, is crucial in permitting formation of the hypernucleosome [64]. Bulkier residues at this position interfere with multimerization [64]. Lastly, we were able to identify several potential interacting residues between layers of dimers that may be important for stabilizing the multimeric complex. These favorable interactions between histone dimers i and $i+2$ and $i+3$, here termed stacking interactions, will contribute to stability of the compacted hypernucleosome. The HMfB hypernucleosome crystal structure shows three stacking interactions, hydrogen bonds from K30 to E61, E34 to R65, and R48 to D14.

Caption Table 1 (page 79)

Dimer–dimer interactions in the tetrameric interface are expected to be essential for hypernucleosome formation. Absence of bulky residues in the first loop and a high number of potential hydrogen bonds in the stacking interface will enhance the compactness and stability of the hypernucleosome. Likely, uncertain and unlikely stacking ability is indicated with +, ±, and –, respectively.

^a Dimer–dimer interface includes residues at positions 46, 49, 59, and 62.

^b Stacking interface includes residues at positions 15–17.

^c For all potential stacking interactions, residue numbering of HMfB was used according to the alignment

Table 1. Hypernucleosome propensity of archaeal histones

	Dimer-dimer interface ^a	Stacking interface ^b	Potential stacking interactions ^c	Hypernucleosome formation	Histone features
Heimdall	LC_3_HA	+	3 (E14-R48, K26-E57, R41-E45)	±	N-terminal tail
Heimdall	LC_3_HB	+	5 (E30-K61, Q14-R48, R13-Q18, K27-E57, K37-E45)	+	
Heimdall	LC_3_HC	+	3 (N34-R65, T15-K41, Y14-Q53)	±	
Loki	GC14_75 (HLKE)	+	2 (D14-R48, K34-E45)	-	Truncated C-term
Loki	CR_4	+	2 (Q14-D48, Q41-Q41)	+	
Odin	LCB_4	-	3 (K30-Q61, K14-E18, E38-R41)	±	
Thor	SMTZ1-45	+	5 (Q30-D61, E34-K65, K14-E48, E37-R41, E26-K58)	+	
Woese	CG1_02_33_12	+	4 (R14-T48, R34-E61, E26-K57, E37-R41)	+	
Pace	CG1_02_31_27	+	4 (S30-K61, E34-K65, K14-T48, E37-K45)	+	
Huber	CG_4_9_14_3_um_filter_31_125 (HA)	+	2 (E14-K48, E14-K18)	±	
Huber	CG_4_9_14_3_um_filter_31_125 (HB)	+	2 (N30-R61, K34-E65)	+	N-terminal tail
Diaphero	CG_4_10_14_0_2_um_filter_31_5	±	4 (E33-R48, E37-R41, E37-R48, E27-K61)	±	
Aenigm	CG1_02_38_14	+	5 (E30-K61, D34-R65, E14-H48, A15-K41, E37-K41)	+	
Micr	<i>M. acidiphilum</i> ARMAN-2	+	3 (E30-K61, K34-Q65, Y2-K48)	+	
Nanohalo	<i>Haloredivivus</i> sp. G17	-	2 (E27-R61, Q37-E45)	-	Truncated N-term
Nanohalo	<i>Nanosalina</i> sp. J07AB43 (HA)	+	5 (Q30-K61, D34-R65, K14-E48, K14-E18, Q37-Q45)	+	
Nanohalo	<i>Nanosalina</i> sp. J07AB43 (HB)	-	2 (Q30-R61, D14-K18)	-	
Nano	<i>N. equitans</i> Kin4-M	+	4 (E30-R61, Q14-K48, Q14(bb)-R41, K37-E45)	+	
Kor	ARK-16 (HA)	±	7 (R30-E61, Q14-K48, K15-E41, E26-R58, E27-R57, E33-K48, R38-E41)	±	
Kor	ARK-16 (HB)	+	3 (Y30-E61, D14-K48, R27-E61)	+	
Thaum	<i>N. gargensis</i> Ga9.2	+	4 (E34-K65, K14-E18, E27-R61, E37-K41)	+	
Bathy	B23	±	4 (R14-V44, E34-K61, E37-R41, E26-R58)	±	N-terminal tail
Bathy	B24	+	3 (E34-R65, K14-E18, E27-R61)	±	Truncated N-term
Bathy	SMTZ-90	+	3 (E34-K65, K41-E45, E27-R61)	+	
Cren	<i>C. mequilingensis</i> IC-167	+	4 (D30-K61, N34-R65, K14-E18, Y37-K48)	+	
Cren	<i>T. pendens</i> Htk5	+	4 (E30-K61, S14-R48, R37-E45, R13-E18)	+	
Cren	<i>V. distributa</i> DSM14429	+	4 (D30-K61, Y34-R65, K14(bb)-R48, K14-E18)	+	
Eury	<i>M. wolnii</i> (HA)	±	4 (N30-E61, E34-K65, E14-K48, K41-E45)	±	
Eury	<i>M. wolnii</i> (HB)	+	5 (E30-K61, E34-K65, N14-R48, N14-Q18, Q41-Q41)	±	
Eury	<i>M. jannaschii</i> DSM2661	-	4 (N30-K61, Q14-R48, K37-Q45, D26-R58)	±	C-terminal tail
Eury	<i>M. methylutens</i>	-	2 (D30-K61, S14-E18)	-	
Eury	<i>T. kodakarensis</i> KOD1 (HTKb)	+	4 (E30-K61, E34-K65, K14-Q48, K26-E58)	+	
Eury	<i>M. fervidus</i> DSM2088 (HMB)	+	3 (K30-E61, E34-R65, D14-R48)	+	

Scrutiny of the modelled archaeal histone dimers reveals that the majority meets our criteria for hypernucleosome formation (Table 1, marked +). We identified two to seven potential stacking interactions for this group of histones, where a higher number of interactions may increase hypernucleosome stability and compactness. Fewer interactions may allow for more “breathing” of the hypernucleosome structure, yielding hypernucleosomes that are more flexible or “floppy”. We predict such structures to be formed also by a number of archaeal histones that do not fully meet our criteria (Table 1, marked ±). For example, *Candidatus Heimdallarchaeota* LC_3 HA and *Candidatus Lokiarchaeota* GC14_75 HLkE have H49N and D59S substitutions, respectively, which likely weakens the crucial hydrogen-bonding interaction at the dimer–dimer interface [105]. Similarly, substitution of the hydrophobic residues 46 and 62 for more hydrophilic or bulkier ones would lead to a less stable dimer–dimer interface, as for *Candidatus Heimdallarchaeota* LC_3 HC and *Candidatus Bathyarchaeota* B23. In the presence of the canonical dimer–dimer interface, bulky substitutions at position 16 likely also result in a more open hypernucleosome structure, as for *Candidatus Odinarchaeota* LCB_4. Three archaeal histone species fail multiple criteria in our analysis, indicating that these cannot form hypernucleosomes. These histone species are *Haloredivivus* sp G17, *Nanosalina* J07AB43 HB, and *Methanococcoides methylutens* that all combine defects in the dimer interface with a bulky substitution at position 16 and few potential stacking interactions (Table 1, marked –). In particular, *Nanosalina* J07AB43 Histone B (HB) shows a H49D substitution and a glutamic acid at position 62, making the dimer surface highly negatively charged and thus very unlikely to interact with another dimer. It is remarkable that most of the histones having N- or C-terminal tails or N- or C-terminal truncations additionally have substitutions in the dimer–dimer and/or stacking interface that will affect hypernucleosome formation. Histones with reduced ability to form compact hypernucleosomes are expected to exhibit different roles in shaping the genome, like DNA bending or site-specific interference with histone multimerization. Interestingly, the genomes of several organisms encode histones that we predict are able to multimerize as well as histones that probably do not multimerize. This suggests that differential expression of these histones could play an important role in regulation of (local) genome compaction and thereby gene regulation.

ARCHAEAL HISTONE TAILS

Prompted by the importance of the eukaryotic histone tails in modulating chromatin structure and function, we constructed a molecular model of a hypernucleosome formed by Histone A (HA) from *Heimdallarchaeota* LC_3 by multitemplate modeling using the HMfB dimer in the hypernucleosome as a structural template for the histone fold and eukaryotic histone H3 and H4 as structural templates for the N-terminal tails (Fig. 5). The model illustrates how three subsequent arginines (R17–R19) could facilitate passing of the tails through the DNA gyres. The tails exit the hypernucleosome through DNA minor grooves, similar to eukaryotic histone tails, and might position their lysine side chains to bind to the hypernucleosomal DNA or to other DNA close by, facilitating (long-range) genomic interactions *in trans*. Like the H4 tail that is subject to acetylation of lysines K5, K8, K12, and K16, lysines in the *Heimdallarchaeal* histone tail may well be subject to acetylation.

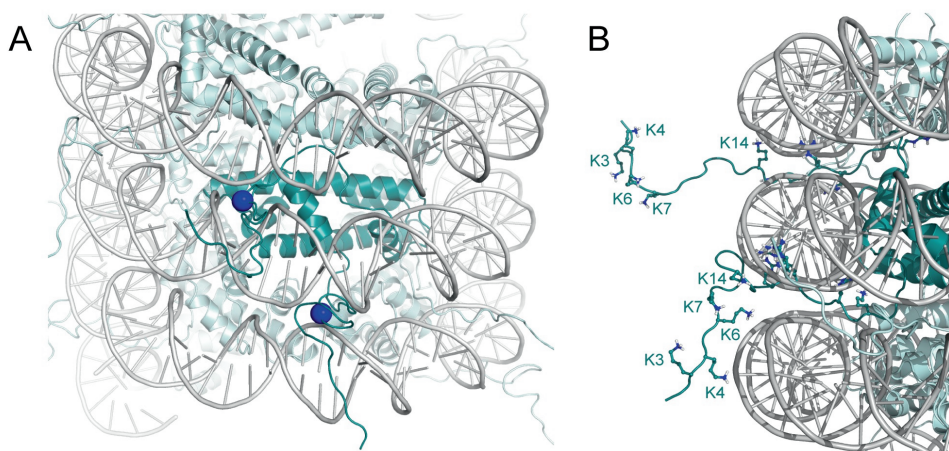


Figure 5. Structural model of a hypernucleosome formed by *Heimdallarchaeota* LC_3 histone HA with N-terminal tails. (A) View showing histone tails protruding through the DNA minor grooves. The R17 C α -atom is shown as a blue sphere to mark the exit point of the tail. (B) Close up of the histone tails with lysine and arginine residues shown as sticks, and N-terminal lysines are labeled. Homodimers of *Heimdall* LC_3 histone HA are shown in teal; one dimer is highlighted in darker colors. Models are based on the structure of HMfB (PDB entry 5T5K); the tail in the top (bottom) of panel B is modeled in the H3 (H4) tail conformation (PDB entry 1KX5).

HMfA VS. HMfB ANALYSIS

HMfA and HMfB from *Methanothermobacter thermautotrophicus* share 84% sequence identity and can form homo- and heterodimers that are homologous to eukaryotic H3H4 dimers. However, as their expression patterns differ, they are suggested to have separate functions in expression regulation *in vivo* (12). The basis for this functional distinction lies in the HMf sequences. To obtain a better understanding of the structural consequences of the amino acid changes between HMfA and HMfB, we used molecular modeling to analyze both histones in terms of interacting residues.

The low-resolution HMfB hypernucleosome crystal structure was water-refined to optimize side-chain conformations and a water-refined HMfA hypernucleosome model was built based on this HMfB structure. From these structures we evaluated the dimer-dimer interface and the stacking interactions that likely stabilize the HMf multimer.

Both HMfA and HMfB form the canonical 4HB: the salt bridge between H49 and D59 as well as hydrophobic residues L46 and L62 are conserved (Fig. 6). In addition, both form a salt bridge between E42 and either R66 (HMfB) or K65 (HMfA) to better accommodate the positive charge in the C-terminus. Also, the critical glycine residue (G16) in the stacking interface loop is conserved in both histones. For HMfB, interactions between R48-D14, K30-E61 and E34-R65 could be identified in the stacking interface (Fig. 6). The former two were also found for HMfA. Although D14 is mutated to N and R48 to K, the interaction can still be established. Mutation of R65 to the slightly shorter K leads to disruption of the E34-R65 connection in HMfA. However, an alternative salt bridge can be formed between the conserved E34 with K45. Interestingly, K45 can simultaneously form a salt bridge to E37, whereas in HMfB there is an arginine at this position.

In conclusion, both dimer-dimer and stacking interfaces for HMfA and HMfB have a similar amount and type of hydrogen bonds. Calculating the theoretical total energy of all interface residues (including hydrophobic interactions) suggests that both HMfA interfaces are energetically more stable than the HMfB interfaces. Reduced repulsive forces between the HMfA dimers because of the charge swap at position 37 and the fact that HMfA has one positive charge less in the C-terminal helix are also in favor of a more stable HMfA complex. However, looking at the residue differences individually and the mutants described in the literature, the R64 to valine swap in the C-terminal

helix is in favor of the HMfB multimer being more stable. Valine at this site makes a hydrophobic contact with I31, stabilizing the histone fold and the position of the C-terminal helix (Fig. 6). An arginine at position 64 disrupts this hydrophobic contact and is a much bigger residue, which is partly compensated by the smaller valine at 31, but might still change the position or flexibility of the C-terminus. A more dynamic C-terminal helix could lead to less compact packing in the hypernucleosome. This was indeed observed for the HMfB mutant V64R in papers by the group of John Reeve (10,21,22), showing reduced electrophoretic mobility of the histone-DNA complex on agarose gel and slightly impaired nucleosome formation (based on EMSA and the amount of supercoiling introduced by the histones). However, whether the stacking interface energy or this C-terminal effect dominates in the formation of a hypernucleosome is not clear from modeling alone.

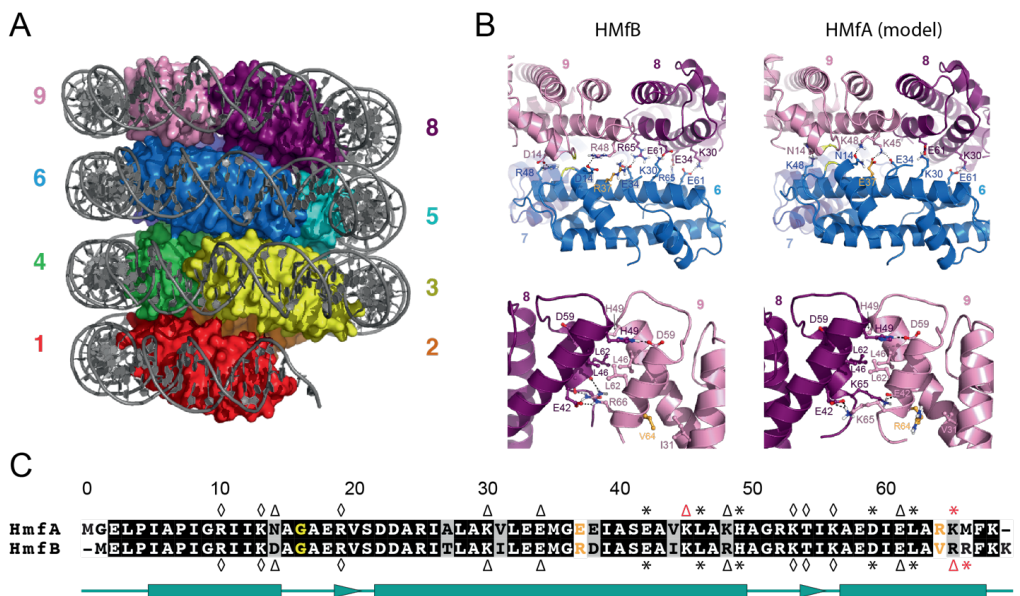


Figure 6. The hypernucleosome and *Methanothermobacter ferredoxin* histones HMfB and HMfA. (A) The HMfB hypernucleosome structure as described by Mattioli *et al.* (PDB: 5T5K). (B) Hypernucleosome interfaces of HMfB (left; water-refined and starting from the 5T5K structure) and HMfA (right; water-refined and starting from the homology model). Top panels show the stacking interface, bottom panels show the tetramer interface. Interacting residues are labeled; hydrogen bonds are shown by dashed lines. Important residue differences between HMfA and HMfB are colored in orange. The closely packed residues G16 in the stacking interface are

shown in yellow spheres ($C\alpha$ only). (C) Sequence alignment of HMfA and HMfB. Conserved residues are shown in black boxes, conservative changes in grey and non-conservative changes in white boxes. Residues are numbered according to HMfB sequence. The two sequences are 84% identical and have 94% similarity. Residues directly contacting the DNA are indicated with diamonds (\diamond), residues forming the tetramerization interface are marked with asterisks (*) and stacking residues are indicated with triangles (\triangle) for HMfA (top) and HMfB (bottom). Differences in interacting residues between HMfA and HMfB are indicated with red symbols. Important residue differences between HMfA and HMfB are colored in orange and G16 residues are shown in yellow, matching panel B.

Biophysical experiments were performed by the labs of John van Noort and Remus Dame on HMfA and HMfB hypernucleosomes to determine their structural and mechanical properties. TPM data and force experiments show that both HMfA and HMfB homodimers enable cooperative compaction of a DNA tether in a left-handed manner (Fig. 7). The extent of this compaction cannot be explained by DNA bending induced by individual dimers and therefore suggests adjacent packing of HMfA (and HMfB) proteins that together wrap DNA, thus indicative of hypernucleosome formation. Force extension curves show two transitions during unfolding of these structures, likely corresponding to unstacking of dimers, followed by unwrapping of the DNA from the dimers (Fig. 7C, D and E). TPM experiments with an HMfA stacking mutant (HMfA_{K31A/E35A}) showed a slightly less compacted structure at saturation levels, less cooperativity and a much higher concentration of HMfA needed to reach saturation (Fig. 7B). This confirms that the stacking interactions are important in cooperative dimer deposition and stabilization of the hypernucleosome, but not crucial for its formation. Interestingly, the biophysical data reveal that HMfB dimers show higher cooperativity during formation of hypernucleosomes and also reach a slightly higher level of compaction (Fig. 7A). In addition, HMfB dimers were found to require higher unstacking and unwrapping forces and to reach a higher dimer occupancy on the DNA tether (Fig. 7D). This suggests that HMfB forms more stable hypernucleosomes than the HMfA, which suggests that the more rigid C-terminal helix in HMfB might indeed be the dominant factor for hypernucleosome stability.

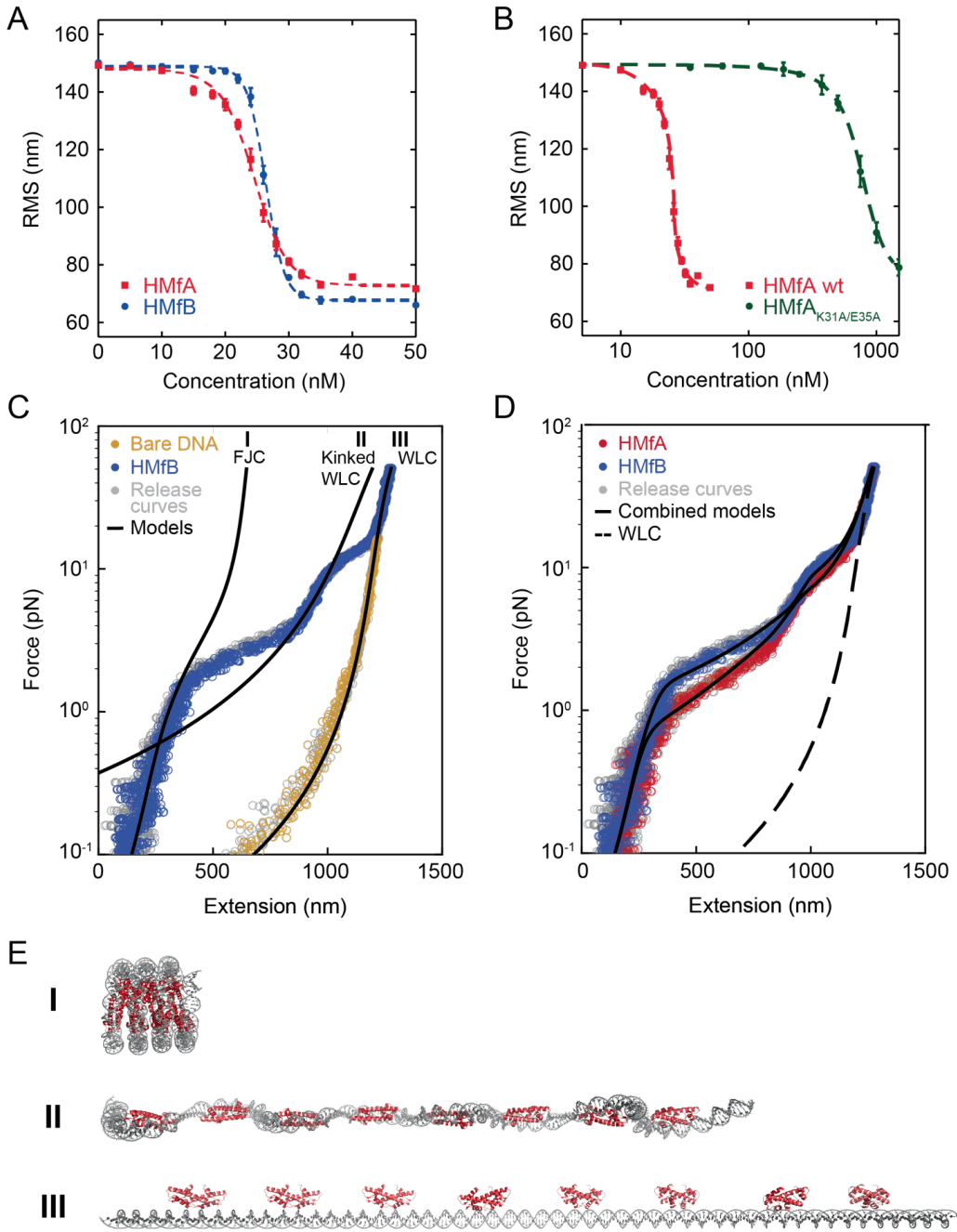


Figure 7. Biophysical experiments on the structural and mechanical properties of HMfA and HMfB hypernucleosomes. (A) HMfA- and HMfB-induced compaction of DNA measured by tethered particle motion (TPM). HMfB dimers reach a slightly higher cooperativity (slope of the curve) and a slightly higher degree of compaction (final RMS). (B) DNA compaction by wild

type HMfA and an HMfA stacking mutant (K31A/E35A) by TPM. HMfA_{K31A/E35A} reaches a slightly less compacted structure, shows less cooperativity and a much higher concentration of HMfA is needed to reach saturation. (C) Force spectroscopy on a saturated HMfB-DNA complex reveals three levels of compaction. These are fitted to a freely-jointed chain (FJC) (I), kinked worm-like chain (WLC) (II), and WLC (III). The transition from regime I to II represents unstacking of the dimers in the hypernucleosome, and the transition from II to II represents the unwrapping of DNA from the dimers. The release curves overlap with the stretch curves, which indicates that the stretch-release cycle was in equilibrium. (D) Comparison between HMfA and HMfB hypernucleosomes. The WLC model for bare DNA was added as a reference (dashed black line). The black lines represent the fitting of the combined unfolding models (as shown in C). (E) Cartoon representation of the three states of the HMfB-DNA complex corresponding to the fits in C, illustrating the 7-fold compaction of the hypernucleosome compared to bare DNA.

DISCUSSION

In this chapter, we used computational modeling to obtain insights into the architectural principles of both archaeal and eukaryotic histone complexes. Building the HMfB hypernucleosome model revealed that the shape and symmetry of the histone dimer determines which nucleosome-like complex can be formed. The shape of the histone surface determines the curvature of the DNA superhelix. If the DNA does not match this curvature, several histone-DNA contacts are lost, destabilizing the complex. A symmetrical histone dimer unit is crucial for formation of a hypernucleosome, as this ensures that all 4HB interfaces are identical and multimerization can continue with constant pitch and bend angle of the superhelix. Consequently, the asymmetric eukaryotic histone dimers are not able to multimerize on DNA. The H3H3 and H4H2B interfaces have different interface residues that lead to non-identical dimer positioning, with an H3H3-like tetramer being flatter and an H4H2B-like tetramer being wider. Moreover, the α N-helix of H3 as well as the docking domain of H2A and the α C helix of H2B block the stacking of subsequent dimers on the nucleosome surface. The nucleosomal DNA has adapted itself to these specific histone dimer shapes, leading to an irregular conformation of the superhelix. Interestingly, HMfB dimers also display a less positively charged surface compared to the very basic eukaryotic histone octamer core (16), making it more likely for the next dimer to come in close proximity of the growing multimer.

A simple homology modeling approach was used to identify other archaeal histones that might be able to form hypernucleosomes, using their primary sequence in alignment with HMfB and histone H4. This showed that the majority of the investigated archaeal histones are likely to be able to form hypernucleosome-like structures, although some may be more flexible or 'floppy' than others. Notably, although it is unlikely that the whole genome of these archaea is wrapped into large hypernucleosomes, we envision some sort of dynamic regulation of compaction by hypernucleosome formation. The presence of multiple histone sequences per species with various hypernucleosome propensity could facilitate differential compaction and gene regulation in archaea.

Surprisingly, some of the histones contain N- or C-terminal tails, much like eukaryotic nucleosomes. Multi-template modeling of one of these showed that the histone tail could adapt a conformation very similar to H3 or H4, with arginine residues allowing for passing of the tail between the DNA gyres.

The presence of lysine-rich histone tails suggests that they might also be post-translationally modified, like eukaryotic histones. Indeed, archaeal genomes are known to have several candidate lysine acetyltransferase and deacetylase enzymes, including proteins from the ELP3 superfamily, to which transcription elongation factor and histone acetyltransferase ELP3 belongs (23-25). Searches using the ProSite database (<http://prosite.expasy.org>, (26)) and Protein Information Resource (<http://pir.georgetown.edu>, (27)) further reveal that the Heimdallarchaeota LC_3 genome contains multiple gene products containing the Gcn5-related N-acetyltransferase domain, which is present in many histone acetyltransferases (28). Interestingly, a potential "reader" protein that binds modified lysines can also be identified. This protein, HeimC3_47440, contains a YEATS-domain, which has recently been shown to bind histone tails that carry acetylated or crotonylated lysines (29-32). Comparison with the closest homolog of known 3D structure, YEATS2 (35% identity, PDB entry: 5IQL, (33)), shows that the binding site for the modified lysine side chain is strictly conserved in the archaeal protein. Notably, only Candidatus Bathyarchaeota, which also features tailed histones, contains a detectable homolog of HeimC3_47440. The presence of lysine-containing N-terminal tails in combination with histone modification writers and readers suggests that archaea use post-translational modifications in a similar way to Eukaryotes as modulators of genome compaction and

gene activity. The tail of the Huberarchaea histone also contains lysine residues that are found at the same position as some of the lysines of the H4 tail. However, no proteins involved in post-translational modification of histone tails have been identified in this phylum.

To further explore the idea of gene regulation by differential expression of different histones within one species, we studied the two histones from *M. fervidus*: HMfA and HMfB. These two histones are known to be expressed in different ratios as a function of growth phase (12), suggesting they have distinct functions that have their basis in the histone sequences and structure. Comparing an HMfA hypernucleosome model to the HMfB hypernucleosome structure showed that despite several amino acid changes, both the 4HB and the stacking interfaces can form salt bridges, H-bonds and hydrophobic interactions that are very similar in number and geometry. Theoretical total energy calculations suggested that the HMfA hypernucleosome might be slightly more stable, although the arginine at position 64 in HMfA might make the C-terminal helix more dynamic, possibly leading to a less compact complex. Biophysical experiments showed that stacking interactions, although not strongly conserved among archaea, indeed play an important role in cooperativity of hypernucleosome formation and stability of the complex. However, the data also showed that HMfB hypernucleosomes are more stable, indicating that the dynamics of the C-terminal helix might be crucial to determine the degree of compaction that can be achieved. This difference in DNA compaction by HMfA and HMfB, which could be further tuned by the formation of heterodimers, might be exploited by the cell as a way of gene regulation in different growth phases. Higher expression levels of HMfA would lead to a less compact hypernucleosome structure and might limit the size of the hypernucleosomal multimer, facilitating access of protein factors involved in further decompaction of chromatin structure and recruitment of transcription or replication factors. In general, we propose that the expression patterns of histone variants within one archaeal species may alter the composition and size of hypernucleosomes, providing a way of gene regulation.

CONCLUSION

In conclusion, we showed that computational modeling of unknown histone structures based on known homologous structures can give insight in their propensity to form multimeric complexes and therefore also in their biological function. Both interacting surface residues and dynamic properties of histone dimers are likely to determine multimerization state, stability and the level of compaction that can be achieved. We propose that different archaeal histone dimers in the same organism can specifically promote or disturb hypernucleosome formation and that this might be functional in gene regulation. Differentiation of histones into asymmetric dimers and the emergence of modifiable N-terminal tails may have led to the evolution of eukaryotic octameric nucleosomes and an epigenetic code providing extensive regulation of DNA-templated processes.

REFERENCES

1. Cox, C.J., Foster, P.G., Hirt, R.P., Harris, S.R. and Embley, T.M. (2008) The archaeobacterial origin of eukaryotes. *Proc Natl Acad Sci U S A*, **105**, 20356-20361.
2. Eme, L., Spang, A., Lombard, J., Stairs, C.W. and Ettema, T.J.G. (2017) Archaea and the origin of eukaryotes. *Nat Rev Microbiol*, **15**, 711-723.
3. Williams, T.A., Foster, P.G., Cox, C.J. and Embley, T.M. (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, **504**, 231-236.
4. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.
5. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074-1080.
6. Dorigo, B., Schalch, T., Bystricky, K. and Richmond, T.J. (2003) Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol*, **327**, 85-96.
7. White, M.F. and Bell, S.D. (2002) Holding it together: chromatin in the Archaea. *Trends Genet*, **18**, 621-626.
8. Sandman, K. and Reeve, J.N. (2005) Archaeal chromatin proteins: different structures but common function? *Curr Opin Microbiol*, **8**, 656-661.
9. Sandman, K., Krzycki, J.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc Natl Acad Sci U S A*, **87**, 5788-5791.
10. Decanniere, K., Babu, A.M., Sandman, K., Reeve, J.N. and Heinemann, U. (2000) Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J Mol Biol*, **303**, 35-47.
11. Starich, M.R., Sandman, K., Reeve, J.N. and Summers, M.F. (1996) NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. *J Mol Biol*, **255**, 187-203.
12. Sandman, K., Grayling, R.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1994) Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus*. *Proc Natl Acad Sci U S A*, **91**, 12624-12628.
13. Bailey, K.A., Chow, C.S. and Reeve, J.N. (1999) Histone stoichiometry and DNA circularization in archaeal nucleosomes. *Nucleic Acids Res*, **27**, 532-536.
14. Pereira, S.L., Grayling, R.A., Lurz, R. and Reeve, J.N. (1997) Archaeal nucleosomes. *Proc Natl Acad Sci U S A*, **94**, 12633-12637.
15. Maruyama, H., Harwood, J.C., Moore, K.M., Paszkiewicz, K., Durley, S.C., Fukushima, H., Atomi, H., Takeyasu, K. and Kent, N.A. (2013) An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Rep*, **14**, 711-717.

16. Mattioli, F., Bhattacharyya, S., Dyer, P.N., White, A.E., Sandman, K., Burkhart, B.W., Byrne, K.R., Lee, T., Ahn, N.G., Santangelo, T.J. *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science*, **357**, 609-612.
17. Spang, A., Caceres, E.F. and Ettema, T.J.G. (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science*, **357**.
18. Ong, M.S., Richmond, T.J. and Davey, C.A. (2007) DNA stretching and extreme kinking in the nucleosome core. *J Mol Biol*, **368**, 1067-1074.
19. Richmond, T.J. and Davey, C.A. (2003) The structure of DNA in the nucleosome core. *Nature*, **423**, 145-150.
20. van Zundert, G.C.P., Rodrigues, J., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J. and Bonvin, A. (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol*, **428**, 720-725.
21. Soares, D.J., Sandman, K. and Reeve, J.N. (2000) Mutational analysis of archaeal histone-DNA interactions. *J Mol Biol*, **297**, 39-47.
22. Sandman, K. and Reeve, J.N. (2001) Chromosome packaging by archaeal histones. *Adv Appl Microbiol*, **50**, 75-99.
23. Reeve, J.N. (2003) Archaeal chromatin and transcription. *Mol Microbiol*, **48**, 587-598.
24. Eichler, J. and Adams, M.W. (2005) Posttranslational protein modification in Archaea. *Microbiol Mol Biol Rev*, **69**, 393-425.
25. Beltrao, P., Bork, P., Krogan, N.J. and van Noort, V. (2013) Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol*, **9**, 714.
26. Sigrist, C.J., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A. and Hulo, N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res*, **38**, D161-166.
27. Wu, C.H., Yeh, L.S., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z., Kourtesis, P., Ledley, R.S., Suzek, B.E. *et al.* (2003) The Protein Information Resource. *Nucleic Acids Res*, **31**, 345-347.
28. Berger, S.L. (1999) Gene activation by histone and factor acetyltransferases. *Curr Opin Cell Biol*, **11**, 336-341.
29. Andrews, F.H., Shinsky, S.A., Shanle, E.K., Bridgers, J.B., Gest, A., Tsun, I.K., Krajewski, K., Shi, X., Strahl, B.D. and Kutateladze, T.G. (2016) The Taf14 YEATS domain is a reader of histone crotonylation. *Nat Chem Biol*, **12**, 396-398.
30. Li, Y., Sabari, B.R., Panchenko, T., Wen, H., Zhao, D., Guan, H., Wan, L., Huang, H., Tang, Z., Zhao, Y. *et al.* (2016) Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol Cell*, **62**, 181-193.
31. Shanle, E.K., Andrews, F.H., Meriesh, H., McDaniel, S.L., Dronamraju, R., DiFiore, J.V., Jha, D., Wozniak, G.G., Bridgers, J.B., Kerschner, J.L. *et al.* (2015) Association of Taf14 with acetylated histone H3 directs gene transcription and the DNA damage response. *Genes Dev*, **29**, 1795-1800.

32. Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y., Li, W. *et al.* (2014) AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell*, **159**, 558-571.
33. Zhao, D., Guan, H., Zhao, S., Mi, W., Wen, H., Li, Y., Zhao, Y., Allis, C.D., Shi, X. and Li, H. (2016) YEATS2 is a selective histone crotonylation reader. *Cell Res*, **26**, 629-632.

3

RAMIFIED ROLLING CIRCLE
AMPLIFICATION FOR SYNTHESIS OF
NUCLEOSOMAL DNA SEQUENCES

Based on the research article: **C.L. van Emmerik**, I. Gachulincova, V.R. Lobbia, M.A. Daniëls, H.A. Heus, A. Soufi, F.H.T. Nelissen and H. van Ingen (2020) Ramified rolling circle amplification for synthesis of nucleosomal DNA sequences. *Analytical Biochemistry*, **588**, 113469

Nucleosome reconstitutions with LIN28B DNA and mobility shift assays with Oct4 were performed by Ivana Gachulincova at the MRC Centre for Regenerative Medicine, University of Edinburgh, United Kingdom.

ABSTRACT

Nucleosomes are a crucial platform for the recruitment and assembly of protein complexes that process the DNA. Mechanistic and structural *in vitro* studies typically rely on recombinant nucleosomes that are reconstituted using artificial, strong-positioning DNA sequences. To facilitate such studies on native, genomic nucleosomes, there is a need for methods to produce any desired DNA sequence in an efficient manner. The current methods either do not offer much flexibility in choice of sequence or are less efficient in yield and labor. Here, we show that ramified rolling circle amplification (RCA) can be used to produce milligram amounts of a genomic nucleosomal DNA fragment in a scalable, one-pot reaction overnight. The protocol is efficient and flexible in choice of DNA sequence. It yields 10-fold more product than PCR, and rivals production using plasmids. We demonstrate the approach by producing the genomic DNA from the human *LIN28B* locus and show that it forms functional nucleosomes capable of binding pioneer transcription factor Oct4.

INTRODUCTION

Nucleosomes are the repeating unit of chromatin, protecting genome integrity and regulating DNA-templated processes like transcription, replication and repair. This makes them a crucial scaffold for protein binding and therefore a highly attractive target for biochemical and structural studies. Nucleosomes are usually reconstituted *in vitro* from the individual histones, H2A, H2B, H3 and H4, and a ~150 bp DNA sequence, referred to as nucleosomal DNA. To obtain stable, homogeneous samples, the DNA sequence used, is typically a so-called strong-positioning sequences, such as the Widom 601 sequence (1) or the human α satellite repeat (2). Although these DNA sequences are widely used in the field, there is a concern that the resulting nucleosomes are more stable than native ones and therefore do not accurately reflect the situation *in vivo*. Therefore, there is an increasing interest in the study of nucleosome dynamics and interactions reconstituted from alternative or genomic DNA sequences. Since structural studies by methods such as x-ray crystallography or NMR spectroscopy usually require milligram quantities of nucleosomes, there is a demand for efficient procedures to obtain nucleosomal DNA with any sequence of choice. We here present a tailored protocol to use rolling-circle-amplification to copy a template into a very long double-stranded repeat that upon cleavage results in the desired nucleosomal DNA fragment.

At present, two methods are regularly used to produce nucleosomal DNA in milligram quantities. Firstly, the DNA can be produced from a plasmid containing multiple repeats of the desired sequence (3). The plasmid is transformed into *E. coli* and amplified by culturing the bacteria. The plasmid is then isolated and digested into the individual repeats, and the plasmid backbone is removed from the product sequence by ion exchange chromatography. In our lab, we get an average yield of ~20 mg of product from 3 liters of culture using a 12-mer repeat of the 601 sequence; the Luger lab reported ~40-60 mg from 6 liters of culture using a 24-mer repeat of the α satellite sequence. Alternatively, regular PCR amplification can be employed for nucleosomal DNA synthesis, starting from a plasmid containing a single copy of the desired sequence. This method is best used for small amounts, but can be scaled up, yielding 2-3 mg of pure product from 40 96-well plates in our lab. The plasmid method is usually preferred, as the yield is much higher. However, construction of the template plasmid can be challenging, forming a bottle neck when switching to different sequences.

Although the single repeat plasmid used in the PCR method does allow sequence flexibility, the method is rather labor-intensive, and the yield is low.

A third option to produce DNA is based on rolling circle amplification (RCA). RCA requires a circular template and a polymerase that has a high processivity and strand displacement capacity, resulting in a long single-stranded product containing many repeats that are complementary to the template sequence. An advantage of RCA over PCR is that it is carried out in a one-step, isothermal reaction. RCA was originally developed for single-stranded amplification of so-called 'padlock' probes for specific DNA sequence detection (4-6). Since then, RCA has been used in a wide variety of applications, ranging from single molecule detection methods to the synthesis of DNA nanostructures and materials (for a review see (7)). For large-scale synthesis of single-stranded DNA, the RCA protocol has been extended to include digestion of the long RCA product into monomers by 'cutter hairpins' (8) or by annealing of a complementary digestion splint to form double-stranded restriction sites (9). This procedure has been used to produce 'monoclonal' single-stranded DNA oligonucleotides on a microgram scale (8) and single-stranded DNA aptamers on a multi-milligram scale (9).

RCA can be tuned to produce double-stranded DNA by the addition of a primer complementary to the product strand. This leads to ramification or (hyper)branching and an overall enhanced amplification factor of the reaction (10,11). This technique, also known as cascade RCA or exponential RCA, has been exploited in diagnostic and biosensing assays as well as in sequencing of single cell genomes to improve detection of low abundance nucleic acid targets (10,12).

Here, we show that rRCA is an efficient and flexible method for large-scale production of dsDNA, in particular for nucleosomal DNA synthesis. The protocol introduced here enables the production of milligram amounts of double-stranded DNA in a scalable, one-pot overnight reaction. By proper design of the template sequence, the long double-stranded product can be directly digested into monomers by a dedicated restriction enzyme. We detail the procedure to design and construct the circular template, to efficiently digest the rRCA product and purify the desired nucleosomal DNA. Our procedure yields 2 mg from a 12 mL reaction, a 10-fold increase in yield compared to the same volume of PCR reactions. The major advantage of the method is the flexibility in choice of DNA sequence. We demonstrate the method by producing

genomic DNA from the 3' end of the *LIN28B* locus and show that it forms functional nucleosomes capable of binding pioneer transcription factor Oct4.

MATERIAL AND METHODS

CONSTRUCTION OF THE STARTING PLASMID

The *LIN28B* template sequence, corresponding to the genomic location hg18-chr6: 105,638,004-105,638,165, was obtained as a sequence-verified gBlock® gene fragment (IDT) including a 15 bp extension on both sides that contain either EcoRI or XbaI restriction sites for cloning into pUC19 (Table 1). The ligated pUC19 vector containing the *LIN28B* gene fragment was used to transform *E. coli* (JM109) cells and the plasmid was purified and sequence-verified before further application.

CIRCULAR TEMPLATE SYNTHESIS

The pUC19 plasmid containing the *LIN28B* template sequence was amplified using standard PCR reactions with Pfu DNA polymerase (produced according to ref. (13), construct available upon request). The template strand was purified by binding the biotinylated PCR product to Streptavidin Sepharose High Performance affinity resin (GE Healthcare), purification of the beads from the reaction mixture by repeated washing with Tris buffer (10 mM Tris HCl pH 7.5, 1 mM MgCl₂), and finally eluting the phosphorylated strand with 0.2 M NaOH. The eluent was neutralized by adding an equal volume of 0.2 M HCl and ethanol-precipitated. The template strand was circularized by heat-annealing the ligation splint and subsequently incubating this partial duplex with 200 U Taq DNA ligase (NEB) per 100 pmol of template in Taq DNA ligase buffer for 3h at 45°C. The circular template was then ethanol precipitated, reconstituted in T4 DNA ligase buffer and treated with 12.5 U Exonuclease I and 125 U Exonuclease III (Thermofisher) per 100 pmol of template at 37°C for 1 hour to remove the splint, unligated template and any other remaining single-stranded impurities. The circular templates were isopropanol precipitated and dissolved in deionized water to a concentration of 1 μM before use in ramified RCA.

Table 1. DNA sequences of the *LIN28B* gene fragment and oligonucleotides

Name	Sequence ^a
<i>LIN28B</i> insert	5' GCAT CGAATT <u>CCCGGG</u> GGTATTAACATATCCTCAGTGGTGAGTA TTAACATGGAACCTACTCCAACAATACAGATGCTGAATAAATGTAGT CTAAGTGAAGGAAGAAGGAAAGGTGGGAGCTGCCATCACTCAGAA TTGTCCAGCAGGGATTGTGCAAGCTTGTGAATAAAG CCCGGGTCT <u>AGACTAGA</u> 3'
Starting primer	5' GGGGGTATTAACATATCCTC 3'
Branching primer	5' GGGCTTTATTCCACAAGCTTGC 3'
Ligation splint	5' TGTTAATACCCCCGGGCTTTATTAC 3'

^aMutated residues from the original *LIN28B* sequence are indicated in bold, restriction sites for EcoRI and XbaI for cloning are underlined, extensions not part of the *LIN28B* sequence are in italic.

PHI29 EXPRESSION AND ASSESSMENT OF ACTIVITY

The gene for Phi29 DNA polymerase was amplified by PCR from a stock of Phi29 phages using primers 5'ACCATGGATCCCATATGCCGAGAAAGATGTATAG3' and 5'ACCATGAATTCTCGAGTTATTTGATTGTGAATGTG3 (restriction sites underlined) and cloned into the NdeI and EcoRI sites of pET28a (Novagen) introducing a N-terminal His6-tag. This construct and detailed protocols for expression and purification are available upon request. Purification scheme is based on a previously described protocol for a His-GST-tagged version (14). Briefly, the protein was expressed in either *E. coli* BL21(DE3) cultivated in LB medium at 37°C for 3 hours, or in *E. coli* Rosetta2 cells (Novagen) in auto-induction medium (ZYM-5052) at 27.5°C overnight (15). After cell lysis by sonication in ice-cold buffer A (25 mM Tris-HCl (pH7.5), 0.5% Tween-20, 0.5% Nonidet P-40 substitute, 5% glycerol, 5 mM β-mercaptoethanol, 1 mM EDTA) with 20 µg/mL PMSF and 100 µg/mL lysozyme, the lysate was treated with DNaseI (2 U/mL) for 10 minutes at room temperature in presence 2.5 mM MgCl₂, 0.5 mM CaCl₂. The protein was purified from the supernatant by His-tag purification over a HisTrap HP Sepharose column (5 mL, GE Healthcare) pre-equilibrated in buffer A supplemented with 10 mM imidazole and 300 mM NaCl. Washing and elution were carried out with 25 and 300 mM imidazole, respectively. Polymerase containing fractions were pooled and subjected to another nuclease treatment with DNaseI (2 U/mL), exonuclease I (4 U/mL) and RNaseI (4 U/mL). Following a final His-tag purification, protein containing fractions were pooled and dialyzed overnight against 1

liter of dialysis buffer (50 mM Tris-HCl (pH7.5), 100 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 0.5% Tween-20, 0.5% Nonidet P-40 substitute and 50% glycerol) at 4°C. The Phi29 DNA polymerase isolate was aliquoted and stored at -20°C. Activity of the isolate was determined by comparing product DNA yields relative to a commercial isolate (Epicentre).

RAMIFIED RCA AND DIGESTION

The *LIN28B* circular template (20 nM) was amplified in a large-scale ramified RCA reaction in 1x CutSmart® buffer (NEB) with 200 U of Phi29 DNA polymerase and 0.5 U inorganic pyrophosphatase (NEB) per mL of reaction, 0.5 mM dNTPs, 2 µM of starting and branching primer and 4mM DTT. Circular template and primer were heat-annealed in CutSmart buffer before adding dNTPs and the enzymes. The rRCA reaction was incubated at 30°C for 5h, followed by the addition of 1000 U SmaI (NEB) per mL of reaction and further incubation at 25 °C overnight. The resulting double stranded 162bp *LIN28B* DNA was purified by anion exchange chromatography on a HiTrap Q HP Sepharose column (5 mL, GE Healthcare) and ethanol precipitated.

NUCLEOSOME RECONSTITUTION

rRCA-produced *LIN28B* DNA was mixed with human purified histone octamers at a 1:1.1 DNA:octamer molar ratio in the presence of 2M NaCl and nucleosomes were reconstituted by salt gradient dialysis. The nucleosomes were analyzed on 6% TBE gel (Novex) in 1X TBE at 90 V for 1 hour and visualized by Ethidium Bromide staining. The nucleosome concentration was calculated by quantifying the intensities of nucleosome bands, using free *LIN28B* DNA as the standard. The densitometric analysis of band intensities was performed using Multi-Gauge software (Fujifilm Science lab).

MOBILITY SHIFT ASSAY WITH OCT4

Full-length human Oct4 fused to an N-terminal 6X histidine tag with a thrombin cleavage site was expressed from a pET28b bacterial expression plasmid in *E. Coli* Rosetta (DE3) pLysS cells. The recombinant protein was purified under denaturing conditions over a His-SpinTrap column (GE Healthcare), desalted using a PD SpinTrap G-25 column (GE Healthcare) and concentrated using an Amicon Ultra-0.5 device (MW cut-off 10 kDa). EMSA was performed with increasing amounts of Oct4 to free *LIN28B*-DNA and *LIN28B*-nucleosomes. The free *LIN28B*-DNA (approx. 10 nM) and

LIN28B-nucleosomes (approx. 25 nM) were incubated with recombinant Oct4 protein (12.5, 25, 50, 100, 200 nM) in DNA-binding buffer (10 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, 10 μM ZnCl₂, 1 mM DTT, 10 mM KCl, 0.5 mg/ml BSA, 5% glycerol) at room temperature for 60 min. Free and Oct4 bound DNA were separated on 5% non-denaturing polyacrylamide gel run in 0.5X TBE at 90V for 4 hours. The gel was stained with ethidium bromide and visualized using SynGene G:Box.

RESULTS

To illustrate the efficient production of any nucleosomal DNA sequences of choice by rRCA, we here focus on reconstituting nucleosomes using a sequence from a well-defined genomic locus in human fibroblasts. The *LIN28B* locus on chromosome 6 contains a well-positioned nucleosome that is the binding site for pioneer factors Oct4, Sox2, Klf4 and c-Myc (OSKM) (16) and is crucial in reprogramming and pluripotency (17,18). A 162 bp sequence from the *LIN28B* locus was successfully used for OSKM binding assays by Soufi *et al.* (19), forming an excellent model system to test the functional quality of the nucleosomal DNA generated by ramified RCA.

The approach is outlined in Figure 1. Briefly, a dsDNA sequence of choice is amplified from a storage vector, one of its strands is purified and circularized to yield the ssDNA template for RCA (Fig. 1A). In ramified RCA, this template is amplified using two primers, a starting primer and a branching primer, and Phi29 DNA polymerase. This enzyme is frequently used for RCA because of its high processivity and its strand displacement capacity (20), high 3'-5' exonuclease activity and low error rate (21,22). The starting primer anneals to the circular template and is elongated into a long, single-stranded repeat to which the branching primer can anneal. In regular RCA, only an equimolar amount of starting primer to circular template is needed, but in rRCA both starting and branching primer need to be in excess to template. Combined with strand displacement by Phi29, this results in the synthesis of a very long, branched, double-stranded repeat of the desired sequence. It should be noted that both starting and branching primer contribute to the ramification, as can be seen in Figure 1B. Subsequent digestion and purification yields the final nucleosomal DNA fragment.

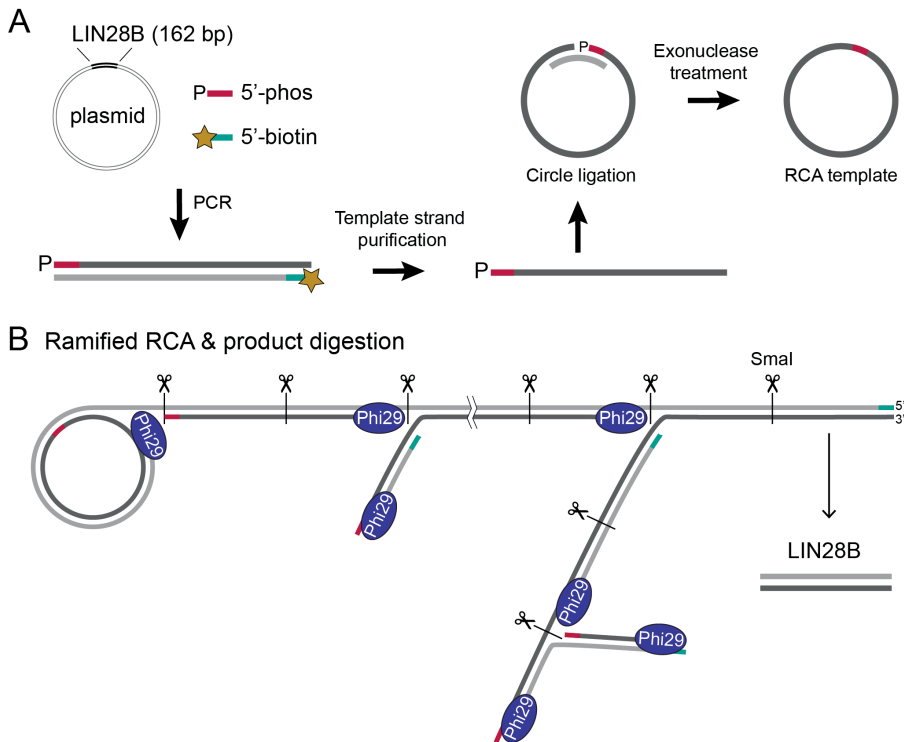


Figure 1. Schematic workflow of ramified RCA for nucleosomal DNA synthesis. (A) Preparation of circular templates, starting from a plasmid with the desired nucleosomal DNA sequence, a 5'-phosphorylated primer and a 5'-biotinylated primer. The template strand (dark gray) is purified by removing the biotin-labeled strand (light gray) using streptavidin beads. The template strand is circularized using a splint to anneal the 5'- and 3'-end together for ligation. (B) Ramified RCA starting from the circular template, two primers (pink and teal), and Phi29 DNA polymerase produces a long, branched dsDNA product. Digestion is performed in the same reaction volume after addition of restriction enzyme SmaI (scissors) to release the double-stranded *LIN28B* product.

DESIGN AND STORAGE OF THE TEMPLATE

The DNA sequence of choice will typically need minor modification to allow efficient use in rRCA. First, the template sequence needs to be designed to encode a restriction site for a blunt-end nuclease in the rRCA product such that the final nucleosomal DNA fragment can be released. In the most straightforward design, the restriction site is formed from the ends of the linear template, and thus overlaps with the ligation site

(see Fig. 1 and 2). Therefore, accurate ligation of these ends into the circular template is essential as mutations in the restriction site impair digestion of the rRCA product and lower the final yield (see also below). Alternatively, the linear template can be designed in a way that the ligation site does not overlap with the restriction site, although care should be taken not to introduce ligation mistakes that interfere elsewhere, e.g. with nucleosome positioning or protein binding. If ligation is found to be inaccurate, the ligation site can be separated from the template sequence by encoding a digestion site on both the 3'- and the 5'-end of the desired sequence. Since the ligation site is then placed between the two restriction sites, it will be cut out of the rRCA product during digestion.

The choice of restriction enzyme can be guided by the sequence of the terminal bases in the desired nucleosomal DNA. As long as these bases are not part of a functional protein binding site, they can be altered at will since nucleosome positioning is encoded within the central 120 bp of the nucleosomal DNA. For *LIN28B*, we changed 4 residues at the termini of the original sequence to construct the recognition site for blunt-end cutter *Sma*I (CCCGGG), the first half of which (CCC) is at the 3'-end and the second half (GGG) at the 5'-end of the template sequence (Fig. 2A and B). Notably, we found that Phi29 is highly active in CutSmart® buffer when supplemented with 4 mM DTT. Thus, rRCA and digestion can be performed in the same reaction buffer, avoiding an additional buffer-exchange.

The RCA method starts from a circular, single-stranded DNA template. Single-stranded oligonucleotides up to 200 nucleotides are nowadays commercially available and could directly be circularized to use as an RCA template. However, the synthesis yield of such long oligonucleotides is usually limited to a few nanomoles and sequence inhomogeneity at the 5'-end can reduce ligation efficiency or result in incorrect ligation products. This not only lowers the amount of desired circular templates but may also give difficulties with the digestion of the product when the restriction site lies in or near the ligation junction. To create a robust stock of template, we therefore decided to generate the ssDNA template from a synthetic dsDNA gene. Such dsDNA fragments can be synthesized with highest sequence accuracy and can furthermore be stored in a vector, allowing for easy, reliable and low-cost amplification in *E. coli*. The desired ssDNA template can then be derived from this material as described below.

The *LIN28B* sequence was cloned in a pUC19 vector by extending the desired sequence with flanking restriction sites (Fig. 2C). Note that these extensions will not end up in the final nucleosomal DNA fragment. This sequence was ordered as a sequence-verified gBlock® double-stranded gene fragment (IDT). The final cloned constructs were verified by sequencing. This simple and straightforward cloning procedure can easily be performed with any other desired sequence and provides in a robust template for the preparation of rRCA single-stranded circles.

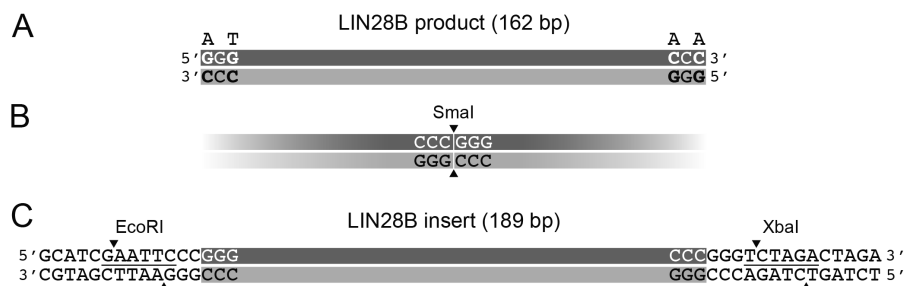


Figure 2. Template design of the *LIN28B* sequence for use in ramified RCA. (A) Four bases of the original *LIN28B* sequence were mutated (in bold, original base indicated on top) to create a SmaI restriction site in the multimeric RCA product (B). (C) The extended *LIN28B* insert contains two restriction sites (EcoRI and XbaI, underlined) to enable cloning into the pUC19 vector.

PREPARATION OF THE TEMPLATE CIRCLES

To obtain the ssDNA template from the storage plasmid, a PCR using two 5'-modified primers is required (Fig. 1A). The primer producing the template strand is 5'-phosphorylated, as the phosphate-group is necessary for ligation to the 3'-end. In order to purify this strand, the primer producing the other strand is 5'-biotinylated. Notably, either strand of a dsDNA molecule of interest can be chosen as template, as both will yield the same double-stranded final RCA product. By binding the PCR product to streptavidin beads and eluting the phosphorylated template strand with sodium hydroxide, pure single stranded product can be obtained. As the streptavidin beads have high binding capacity, no PCR product purification is necessary to remove excess biotinylated primers before binding to the streptavidin beads. This procedure yielded pure ssDNA *LIN28B* fragment in a straightforward manner (Fig. 3A).

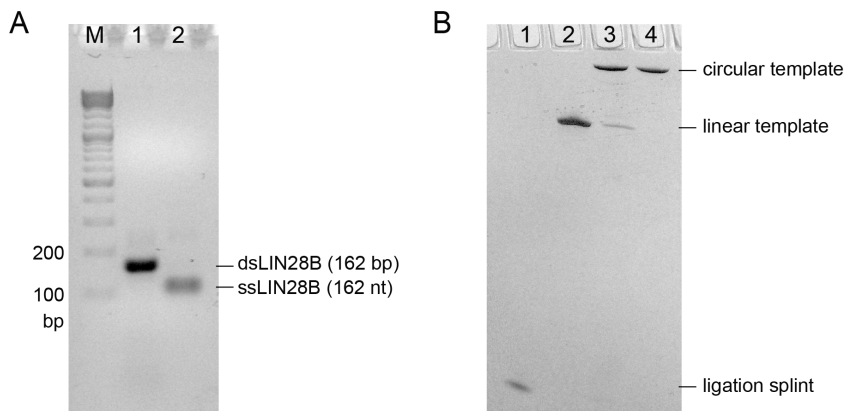


Figure 3. Template strand production and circularization. (A) Agarose gel (2.5%, EtBr staining) showing template strand purification: crude PCR product (lane 1) and purified *LIN28B* strand (lane 2). (B) Denaturing PAGE gel (12%) showing circularization and exonuclease digestion of the template strand: Ligation splint (lane 1, 26 nt), purified template strand (lane 2, 162 nt), ligated template circles (162 nt), crude (lane 3) and after exonuclease clean-up (lane 4).

To circularize the linear template strand a short ssDNA fragment, the ligation splint, is needed to anneal both ends together. This creates a double-stranded region with a nick that will be ligated using a DNA ligase. Although T4 DNA ligase is a very efficient and commonly used enzyme for ligating nicks, it is also capable of ligating the termini if there is a 1 or 2 nucleotide gap present or if the nick contains mismatches (23-25). These gaps and mismatches could occur due to incorrect annealing of both ends on the ligation splint or by heterogeneity in the 5'-end of the primer used in PCR, commonly present in synthetic oligonucleotides. Since inaccurate ligation will cause mutations in the restriction site, special care needs to be taken to avoid this. We therefore used the thermostable Taq DNA ligase, which has a higher ligation accuracy and has no or little activity on gaps or mismatches at the ligation site (26). Taking advantage of its thermostability, the accuracy of splint annealing can be further increased by carrying out the ligation at elevated temperature. For *LIN28B* a 26nt ligation splint was used (see Table 1) in a ligation reaction with Taq DNA ligase at 45°C, resulting in a homogenous pool of circular templates (Fig. 3B).

After ligation, the reaction mix not only contains the circular template but also left-over splint and linear template. These could prime either on the circle or on the product during RCA, resulting in single stranded by-products. To avoid these contaminations

of the product, all linear ssDNA fragments remaining in the ligation are digested with exonuclease I and III. Figure 3B shows this treatment results in pure circular templates. Sequencing analyses using the PCR primers confirmed that the purified circles contain the correct *LIN28B* sequence.

OPTIMIZATION OF RAMIFIED RCA REACTION CONDITIONS

Ramified RCA is performed as a one-pot reaction at constant temperature (30°C for Phi29 DNA polymerase). The optimal reaction buffer will depend on both the polymerase and restriction enzyme used, as mentioned above. For rRCA, the yield depends on the degree of branching and thus on the amount of starting and branching primer. Both primers need to be added in a 1:1 ratio to avoid contamination with ssDNA. To maximize the yield, also the amount of dNTPs and Phi29 DNA polymerase can be increased as these can become limiting. Finally, the purity of the circular template is essential to avoid by-products from remaining linear fragments (see above).

Initial rRCA reactions were performed using template circles that were purified after a standard 30-minute exonuclease treatment. While the templates appeared pure on gel-analysis, considerable ssDNA *LIN28B* product (and multimers) were generated after digestion, indicating presence of residues linear template or annealing splint (Fig. 4A). By extending the exonuclease treatment three-fold, these by-products could be largely eliminated (Fig. 4A). Unexpectedly, in negative controls carried out without starting and branching primers, RCA product formation was still observed when starting from these pure template circles (Fig. 4B). This suggests that there are trace levels of DNA in the reaction mixture that can act as primer. These may originate from the Phi29 DNA polymerase stock, as co-purification of DNA is hard to avoid in preparations of this enzyme due to its high affinity for ssDNA and dsDNA (27). Alternatively, trace levels of DNaseI in the Phi29 stock could generate linear ssDNA fragments capable of priming from the circular template directly.

To determine the optimal primer and Phi29 amounts, a series of 50 μ L reactions were carried out using an excess of dNTPs (500 μ M), and in presence of pyrophosphatase to avoid pyrophosphate-induced sequestering of Mg^{2+} . All reactions showed formation of a large RCA product that remains in the well of the agarose gel (see Fig. 4B and C, top panel). Since the product was difficult to pipet due to its high viscosity, the rRCA

yield was assessed only after digestion by *Sma*I, resulting in *LIN28B* monomers of the expected size (162 bp) together with some dimers (324 bp) and multimers resulting from incomplete digestion (Fig. 4C, bottom panel).

First, the non-ramified control reaction, using only starting primer in equimolar ratio to the circular template, generated double-stranded product of the same size as *LIN28B* monomers, while a long single-stranded product was expected (Fig. 4C, lane 1). This may be caused by trace levels of DNA impurities as mentioned above, possibly in combination with aspecific priming of the starting primer or template-switching by Phi29 (28).

Second, addition of an equimolar amount of the branching primer showed double-stranded product formation with negligible increase in yield compared to the non-ramified reaction (Fig. 4C, lane 2). Under these equimolar conditions, the branching primer can anneal to the product as soon as a full ssDNA *LIN28B* repeat is synthesized, resulting in only one or a few repeats of double-stranded product and no ramification, leaving the majority still single-stranded.

Reactions with either 50- or 100-fold molar excess of both primers, showed significant and progressively increasing yield of double-stranded product, indicating successful ramification (Fig. 4C, lane 3 and 5). This confirms that effective branching only occurs at excess of both primers compared to the template. At these high primer concentrations, the amount of Phi29 DNA polymerase becomes limiting, as seen by the increase in yield upon adding two- or three-fold more Phi29 at constant primer concentration (Fig. 4C, lane 4, 6 and 7). Further increase of either primers or Phi29 may increase yield even more. Since very high excess polymerase is known to reduce polymerization efficiency, we choose to not use even higher excess of polymerase, but rather scale-up the reaction volume (see below).

In all cases, reaction time was limited to five hours, at which point the reaction mixtures turned highly viscous due to product formation. We presumed reaction progress would be significantly reduced and thus little benefit would be gained by extending the reaction for longer.

While these results highlight a sensitivity of Phi29 DNA polymerase to the presence of trace levels of primers, large excess of starting and branching primer under ramified

conditions will ensure correct priming and product formation. The results further demonstrate that maximum rRCA yield of dsDNA product is obtained at high molar excess of both primers in combination with elevated levels of Phi29 DNA polymerase. These conditions are most likely independent of the DNA sequence being produced.

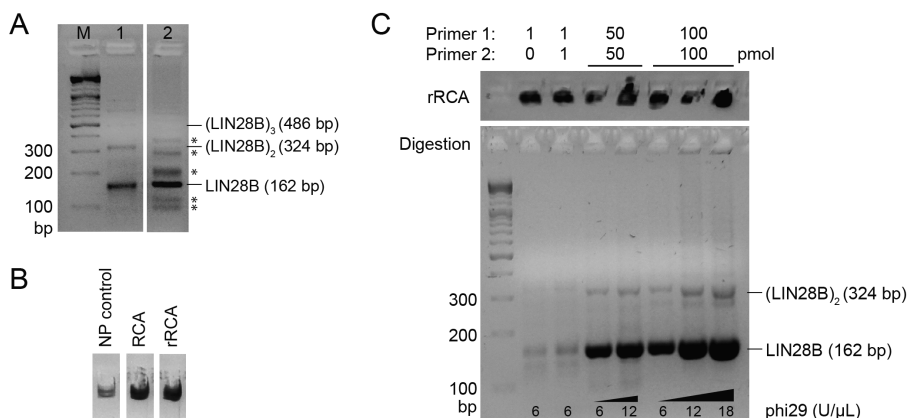


Figure 4. Optimization of ramified RCA reaction conditions. (A) Agarose gel (2.5%, EtBr staining) showing digested rRCA products from a reaction with a clean circular template (lane 1) and with a circular template that still contained impurities (lane 2). Single-stranded by-products are indicated with an asterisk (*). (B) RCA products for ‘no primer’ (NP) control compared to product from RCA (1 pmol starting primer) and rRCA (50 pmol of both primers) on 0.8% agarose gel (EtBr staining). The circular template (1 pmol) used in this experiment was treated with exonuclease I and III for 1.5h. (C) Optimization of primer amounts for ramification and adjustment of Phi29 DNA polymerase amount (2.5% agarose gel, EtBr staining). Top panel shows rRCA product, bottom panel shows *Sma*I-digested product. Before digestion, the rRCA yield is hard to compare, because of inconsistent pipetting due to high viscosity of the sample. Primer amounts are indicated above each lane, Phi29 concentrations are indicated below each lane.

LARGE-SCALE PRODUCTION, DIGESTION & PURIFICATION

For large-scale production of nucleosomal DNA, we take advantage of the fact that both rRCA and digestion are performed in a single reaction tube without the need of thermocycling. This means the reaction can easily be scaled up by increasing the reaction volume and reaction components accordingly. For large-scale production of *LIN28B* DNA, we performed both a 2 mL and a 12 mL synthesis in a single tube under

conditions determined above for 5 hours. In both cases, viscosity of the solution became very high at this point, which was quickly lowered after addition of the restriction enzyme *Sma*I. Digestion of the rRCA product into 162bp *LIN28B* monomers was almost complete after overnight incubation (Fig. 5A). The remaining larger fragments were separated from the main product by anion exchange chromatography using a very shallow salt gradient, resulting in pure nucleosomal DNA (Fig. 5B and C). We obtained 0.35 mg (3500 pmol) of purified product starting from 40 pmol of circular template in a 2 mL reaction volume, which is almost a 10-fold increase in yield compare to the same volume of regular PCR reactions. The 12 mL synthesis yielded 2.0 mg of pure *LIN28B*, demonstrating that the rRCA scales linearly with reaction volume.

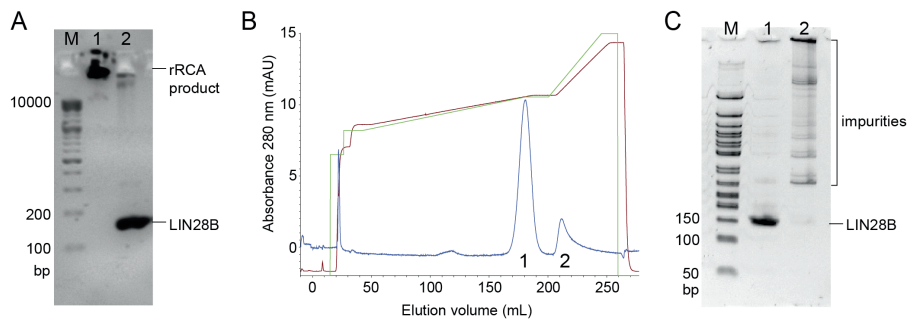


Figure 5. Ramified RCA, *Sma*I digestion and ion exchange chromatography purification. (A) Ramified RCA product (lane 1) and *Sma*I digestion (lane 2) (B) Purification chromatogram of crude digested rRCA product using a 5 mL HiTrap Q HP column; A280 trace in blue, applied salt gradient in green (0-100% B, 1M NaCl), measured conductivity in brown (mS). (C) 5% native PAGE gel showing the pooled fractions of peak 1 and 2, indicated in the chromatogram (panel B).

RECONSTITUTION OF LIN28B NUCLEOSOMES AND OCT4 BINDING

We next aimed to demonstrate that the nucleosomal DNA generated in our ramified RCA approach can be used to reconstitute functional nucleosomes. We used the *LIN28B* DNA from the large-scale production together with recombinantly expressed human histones to reconstitute human nucleosomes by salt-gradient dialysis, according to a previously published method (29). The reconstitution was assessed by native PAGE analysis (Fig. 6A). A clear band shift was observed, indicating nucleosome formation. The efficiency of reconstitution was estimated from gel band

intensities to be ~70%, which is comparable to what was previously published for *LIN28B* (19). We thus conclude that rRCA derived nucleosomal DNA can be used for nucleosome reconstitution in the same manner as DNA from other sources.

To show that these nucleosomes are functional, we assayed their ability to bind pioneer transcription factor Oct4. Oct4 has been reported to associate with *LIN28B* nucleosomes *in vitro* in a sequence specific manner, as shown by DNase I footprinting (19). Addition of recombinant Oct4 protein to our *LIN28B* nucleosomes induced a clear supershift in an EMSA experiment, indicating formation of an Oct4-nucleosome complex (Fig. 6B). Both the supershift and affinities for both free DNA and nucleosomes are in good agreement with previously published results (19). Together, these results show that the rRCA-produced *LIN28B* DNA has the required quality in terms of purity and sequence to enable reconstitution of native nucleosomes capable of binding pioneer transcription factor Oct4.

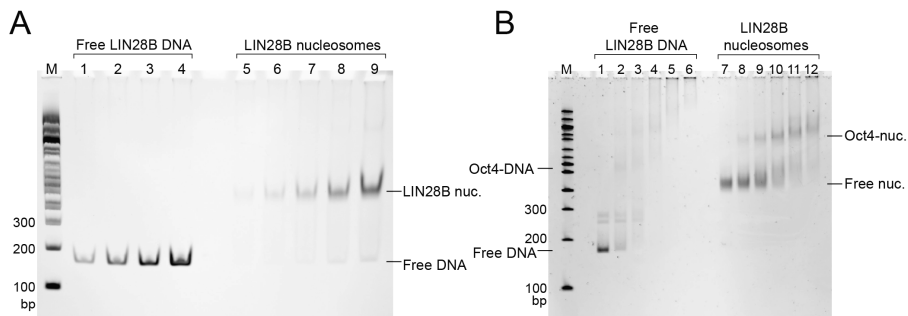


Figure 6. (A) Nucleosome reconstitution with rRCA-produced *LIN28B* DNA (162bp) and human histones. Native PAGE (6%) lanes 1-4 contain 10, 25, 50 and 100 ng of free *LIN28B* DNA used for quantification of the nucleosome bands. DNA was mixed with histone octamer in a 1:1.1 ratio for reconstitution. Lanes 5-9 show increasing amounts of reconstituted nucleosomes estimated to contain 7, 17, 35, 71 or 142 ng of bound DNA, respectively, which is ~70% of the input free DNA. DNA in the nucleosome has approximately a 2-fold lower signal intensity compared to free DNA. (B) EMSA of Oct4 binding to free *LIN28B* DNA (10 nM) and assembled *LIN28B* nucleosomes (25 nM). Native PAGE (5%) lanes 1-6 and lanes 7-12 contain 0, 12.5, 25, 50, 100 and 200 nM of recombinant Oct4.

We here showed that ramified RCA is an efficient and flexible alternative for the large-scale production of nucleosomal DNA in milligram quantities. Compared to plasmid and PCR-based production, our rRCA protocol combines the flexibility in choice of sequence from PCR with the excellent yields of a multi-repeat product from the plasmid-based method. The straightforward design procedure requires only minimal changes to the terminal ends of DNA sequence of choice in order to create the circular template needed for RCA. A crucial aspect in the design of the template is the incorporation of a restriction site to digest the long, branched product. Judicious choice of restriction enzyme allows product digestion in the same reaction volume as the rRCA reaction, boosting the time-efficiency of the method. Production, digestion and purification can be performed in two days, the whole procedure including preparation of circular templates can be completed within a week. The yield is almost 10-fold higher than in the same volume of PCR reactions. Importantly, the one-pot rRCA reaction can easily be scaled up with a linear increase in yield. We here selected a DNA sequence corresponding to a well-positioned nucleosome at the 3' end of the *LIN28B* locus. Since this sequence with 56.8% AT content has no particular features, we expect that the method can generally be applied to most DNA sequences, including larger templates for reconstitution of nucleosomal arrays on genomic DNA. Performance with highly repetitive sequences may be compromised by fidelity problems of the polymerases.

RCA requires a polymerase with extreme processivity and strand displacement, such as provided by Phi29 DNA polymerase. Since large-scale production runs also require high amounts of enzyme, in-house production is advisable to reduce costs (the construct is available upon request). Although this might be a barrier for labs that are not in need of continuous production of nucleosomal DNA, production is efficient (a 1 L culture provides Phi29 for approximately 4000 mL of rRCA) and the resulting Phi29 can be stored long-term at -20°C.

Addition of pyrophosphatase was necessary to avoid the scavenging of Mg^{2+} by the formed pyrophosphate. This may be circumvented by performing the reaction in a dialysis bag in reaction buffer to avoid accumulation of pyrophosphate. Such a setup may improve yield by increasing buffer capacity and avoiding increases in phosphate concentration during the reaction, which could inhibit Phi29. The high viscosity

observed towards the end of the rRCA reaction likely limits the diffusion of polymerase and reactants through the solution, and may thus reduce the potential yield. Performing rRCA and product digestion simultaneously, preventing the formation of very long *LIN28B* repeats, could therefore further improve the yield. However, this requires methylation of the circular templates in combination with a methylation-sensitive restriction enzyme, to avoid digestion of the template.

The rRCA yield was improved by increasing the primer to template ratio, as well as the dNTP and Phi29 concentration. Although rRCA can provide exponential amplification of the template as demonstrated at small-scale for Bst DNA polymerase (30), we did not observe this in our work. Analysis of the yield showed that the starting template was amplified effectively ~100-fold. The amplification here is presumably limited by the high concentration of the reactants and product leading to high viscosity. Dilution of the reaction could alleviate this, but lowered reaction rates may counteract reaching a higher yield. While further improvement in yield may be possible, we demonstrated here that our current protocol is effective and can furthermore be easily scaled. In comparison, exponential amplification in PCR is also only reached at very low template levels, thus requiring a large volume of reactions to reach milligram scale of purified product. In contrast, our rRCA protocol offers a highly practical execution without the need to prepare dozens of PCR-plates, and in particular when access to multiple PCR machines is limited. Moreover, the lack of thermocycling in rRCA leads to more efficient dNTP incorporation as a result of decreased degradation (9). Even if our protocol relies on PCR to generate the template strand, one PCR plate will generate sufficient template for a ~36 mL RCA reaction, which would result in ~6 mg of DNA, equivalent to ~120 plates of PCR.

In conclusion, we presented a rRCA-based method to allow the efficient and flexible large-scale synthesis of nucleosomal DNA sequences, or any other dsDNA of comparable length. The protocol was used to produce multi-milligrams of human, genomic nucleosomal DNA with high purity. We demonstrated that rRCA-produced *LIN28B* DNA can be used to reconstitute stable and functional nucleosomes that are capable of binding pioneer transcription factor Oct4. Due to its ease of use and flexibility in sequence design, we believe this method is an ideal tool to produce a wide variety of nucleosomal DNA sequences to study the structure, dynamics, and interactions of nucleosomes or nucleosomal arrays, in particular of genomic nucleosomes as they occur *in vivo*.

REFERENCES

1. Lowary, P.T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*, **276**, 19-42.
2. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.
3. Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M. and Luger, K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol*, **375**, 23-44.
4. Fire, A. and Xu, S.Q. (1995) Rolling replication of short DNA circles. *Proc Natl Acad Sci U S A*, **92**, 4641-4645.
5. Liu, D., L Daubendiek, S., A Zillman, M., Ryan, K. and Kool, E. (1996) *Rolling Circle DNA Synthesis: Small Circular Oligonucleotides as Efficient Templates for DNA Polymerases*.
6. Baner, J., Nilsson, M., Mendel-Hartvig, M. and Landegren, U. (1998) Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res*, **26**, 5073-5078.
7. Ali, M.M., Li, F., Zhang, Z., Zhang, K., Kang, D.-K., Ankrum, J.A., Le, X.C. and Zhao, W. (2014) Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chemical Society Reviews*, **43**, 3324-3341.
8. Ducani, C., Kaul, C., Moche, M., Shih, W.M. and Högberg, B. (2013) Enzymatic production of 'monoclonal stoichiometric' single-stranded DNA oligonucleotides. *Nature Methods*, **10**, 647.
9. Nelissen, F.H.T., Goossens, E.P.M., Tessari, M. and Heus, H.A. (2015) Enzymatic preparation of multimilligram amounts of pure single-stranded DNA samples for material and analytical sciences. *Analytical Biochemistry*, **475**, 68-73.
10. Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C. and Ward, D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet*, **19**, 225-232.
11. Zhang, D.Y., Brandwein, M., Hsuih, T. and Li, H.B. (2001) Ramification amplification: a novel isothermal DNA amplification method. *Mol Diagn*, **6**, 141-150.
12. Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W. and Church, G.M. (2006) Sequencing genomes from single cells by polymerase cloning. *Nature Biotechnology*, **24**, 680-686.
13. Nelissen, F.H.T., Girard, F.C., Tessari, M., Heus, H.A. and Wijmenga, S.S. (2009) Preparation of selective and segmentally labeled single-stranded DNA for NMR by self-primed PCR and asymmetrical endonuclease double digestion. *Nucleic Acids Research*, **37**, e114-e114.

14. Blainey, P.C. and Quake, S.R. (2010) Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Research*, **39**, e19-e19.
15. Studier, F.W. (2014) In Chen, Y. W. (ed.), *Structural Genomics: General Applications*. Humana Press, Totowa, NJ, pp. 17-32.
16. Soufi, A., Donahue, G. and Zaret, Kenneth S. (2012) Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell*, **151**, 994-1004.
17. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R. *et al.* (2007) Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science*, **318**, 1917.
18. Shyh-Chang, N., Zhu, H., Yvanka de Soysa, T., Shinoda, G., Seligson, Marc T., Tsanov, Kaloyan M., Nguyen, L., Asara, John M., Cantley, Lewis C. and Daley, George Q. (2013) Lin28 Enhances Tissue Repair by Reprogramming Cellular Metabolism. *Cell*, **155**, 778-792.
19. Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M. and Zaret, K.S. (2015) Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, **161**, 555-568.
20. Blanco, L., Bernad, A., Lazaro, J.M., Martin, G., Garmendia, C. and Salas, M. (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem*, **264**, 8935-8940.
21. Esteban, J.A., Salas, M. and Blanco, L. (1993) Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J Biol Chem*, **268**, 2719-2726.
22. Garmendia, C., Bernad, A., Esteban, J.A., Blanco, L. and Salas, M. (1992) The bacteriophage phi 29 DNA polymerase, a proofreading enzyme. *J Biol Chem*, **267**, 2594-2599.
23. Nilsson, S.V. and Magnusson, G. (1982) Sealing of gaps in duplex DNA by T4 DNA ligase. *Nucleic Acids Res*, **10**, 1425-1437.
24. Goffin, C., Bailly, V. and Verly, W.G. (1987) Nicks 3' or 5' to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase. *Nucleic Acids Res*, **15**, 8755-8771.
25. Wu, D.Y. and Wallace, R.B. (1989) Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene*, **76**, 245-254.
26. Barany, F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc Natl Acad Sci U S A*, **88**, 189-193.
27. Takahashi, H., Yamazaki, H., Akanuma, S., Kanahara, H., Saito, T., Chimuro, T., Kobayashi, T., Ohtani, T., Yamamoto, K., Sugiyama, S. *et al.* (2014) Preparation of Phi29 DNA polymerase free of amplifiable DNA using ethidium monoazide, an ultraviolet-free light-emitting diode lamp and trehalose. *PLoS One*, **9**, e82624.

28. Ducani, C., Bernardinelli, G. and Högberg, B. (2014) Rolling circle replication requires single-stranded DNA binding protein to avoid termination and production of double-stranded DNA. *Nucleic Acids Research*, **42**, 10596-10604.
29. Tanaka, Y., Tawaramoto-Sasanuma, M., Kawaguchi, S., Ohta, T., Yoda, K., Kurumizaka, H. and Yokoyama, S. (2004) Expression and purification of recombinant human histones. *Methods*, **33**, 3-11.
30. Beals, T.P., Smith, J.H., Nietupski, R.M. and Lane, D.J. (2010) A mechanism for ramified rolling circle amplification. *BMC Molecular Biology*, **11**, 94.

THE ROLE OF THE N-TERMINAL REGION IN
AUTOREGULATION OF NUCLEOSOME
REMODELING BY ISWI

Assignment of ISWI ATPase domain resonances was performed by Vincenzo Lobbia at the Bijvoet Center for Biomolecular Research, Utrecht University.

ABSTRACT

Chromatin remodelers are an essential class of chromatin factors for proper control of gene expression and protection of genome integrity. Remodelers are ATP-dependent molecular motors that move or evict nucleosomes in a mechanism that relies on the motion of the two ATPase lobes. To prevent spurious consumption of ATP, these lobes are locked in an inactive conformation in the free state of the remodeler, and this autoinhibition is only relieved by interaction with the substrate, in particular the nucleosomal DNA and histone H4 tail. In this study, we applied Fluorescence Anisotropy (FA) and Nuclear Magnetic Resonance (NMR) to identify and characterize conformational changes of the ISWI remodeler upon interaction with its activating nucleosomal epitopes. We show that the ISWI ATPase domain is only able to bind the H4 tail in presence of dsDNA and that this ISWI-DNA-H4 complex requires presence of ATP-mimic ADP-BeFx for stability. We find that the N-terminal region (NTR) of ISWI, which locks the ATPase lobes in their inactive conformation, is not stable as an isolated domain and likely to be bound to the ISWI core at all stages of activation. Furthermore, our data indicate that dsDNA binding alone induces a conformational change in ISWI with an additional change upon H4 tail binding. Based on these data, we propose that the first step of autoinhibition relieve is the interaction of ISWI to dsDNA. In this model, basic residues in the NTR mediate non-specific binding to DNA thereby loosening its binding to the ISWI core and thus allowing for a conformational change of the ATPase lobes. Once ISWI is bound to the nucleosome in proximity of the H4 tail, this tail binds to the second lobe to stabilize the active conformation and promote further relieve of autoinhibition by the NTR.

INTRODUCTION

Chromatin function crucially depends on a variety of factors that interact with nucleosomes to regulate DNA-templated processes like gene expression, DNA repair and DNA replication. One important class of chromatin factors is formed by nucleosome remodelers, part of the helicase-like superfamily 2 (SF2) (1). Within this superfamily, remodelers form the Snf2-like family, which is further subdivided into 24 subfamilies. Major subfamilies of ATP-dependent remodelers include Chd, INO80, SWI/SNF and ISWI. They all contain a common core of two recA-like domains and share a series of seven short sequence motifs, called the helicase motifs (I, Ia, II, III, IV, V and VI), that are involved in DNA and ATP binding (1). The subfamily members contain additional domains for target specificity, autoregulation or interaction with other proteins and usually function in larger complexes with epigenetic readers or modifiers and additional effector proteins. Most remodelers are able to slide nucleosomes along the DNA (2). Some can catalyze additional processes like eviction or exchange of histones for histone variants, *de novo* nucleosome assembly or regular spacing of nucleosomes on the DNA. This local reshaping of the chromatin state can provide access for additional chromatin factors to bind or prepare chromatin for recompaction after transcription, replication or repair. This chapter will focus on ISWI, a well-studied subfamily of ATP-dependent nucleosome remodelers that facilitate nucleosome sliding and spacing.

Within a cellular context, ISWI remodelers are part of NURF, CHRAC and ACF remodeling complexes (2). In vitro, the ATPase domain of ISWI was shown to also be an autonomous remodeling machine able to independently slide nucleosomes, although its accessory domains are required to achieve regular spacing, directionality or regulation of activity (3). These accessory domains are unique to the ISWI subfamily and include the HAND SANT SLIDE (HSS) domain, which is a DNA binding domain that can 'sense' linker DNA length between nucleosomes and controls the directionality of sliding to achieve regular nucleosome spacing (4-6). Two accessory domains are crucial for the auto-inhibition of ATPase activity in absence of a nucleosome, the so-called "negative C-terminal" or NegC region and the N-terminal region (NTR) of ISWI (Fig. 1A) (7,8). The NTR is highly charged and contains several conserved sequence motifs, including a 'post-post-helicase-SANT-associated' (ppHSA) motif present in

several Snf2-family remodelers, a weakly conserved AT-hook, and the so-called AutoN and AcidicN motifs, most of them still of unknown function (8) (Fig. 1A). ISWI autoinhibition is released in presence of the nucleosome, in particular through the histone H4 tail and (extra)nucleosomal DNA (7,9,10). In this chapter we will focus on the relieve of autoinhibition to understand how ISWI reaches its active conformation upon binding of a nucleosome.

Several X-ray and cryoEM structures of ISWI have become available in recent years, capturing remodelers in different conformations, suggesting different stages of activation. The first structure showed MtISWI (ISWI from *Myceliophthora thermophila*) in an inactive conformation (11). The two ATPase lobes in this structure are positioned such that the helicase motifs are separated far apart and thus unable to bind nucleosomal DNA or ATP. This structure showed that the NTR contains several α -helices that are packed against the surface of both ATPase lobes, holding them together and preventing rearrangement for DNA or nucleotide binding (Fig. 1B). The authors suggested that this conformation of the NTR is released in presence of the histone H4 tail, that contains a basic motif (KR₁₈HRK) similar to the AutoN motif, R₁₄₉RRR in MtISWI, HR₉₁HRK in DmISWI (*Drosophila melanogaster* ISWI), that is bound to the second ATPase lobe (Lobe2). Competition of the H4 tail for binding at Lobe2 may thus release the NTR from Lobe2 and thereby allow for a structural rearrangement of the two lobes into the active conformation. A crystal structure of Lobe2 bound to an H4 tail peptide indeed showed that the tail binding is incompatible with binding of the AutoN motif (11). Interestingly, a cross-linking-based study reported AutoN binding at a non-overlapping site in close proximity of the H4 tail (8). Both studies show that the binding modes of AutoN and H4 tail are not identical, arguing against an autoinhibition release mechanism by molecular mimicry and direct competition (8,11). The conformation of NegC, the other autoinhibitory domain of ISWI, was likely affected by crystal packing as it was bound to Lobe2 of an adjacent molecule (11). An integrative modeling approach based on XL-MS, SAXS and computational modeling revealed that, just as the NTR, NegC and the HSS domain are packed against the ATPase domain in solution, thereby caging the catalytic core (12). Binding of the HSS domain to linker DNA of sufficient length may thus be an additional step in the release of autoinhibition.

The various ISWI structures bound to the nucleosome in either nucleotide-free, ADP-bound or ADP-BeFx-bound conditions invariably show a dramatically different conformation of the ATPase lobes, allowing them to bind the nucleosomal DNA (13-15) (Fig. 1B). In this state, the basic motif in the H4 tail occupies the same binding site on Lobe2 as the AutoN motif in the inactive state. Notably, the majority of the NTR as well as NegC are not observed in these structures, suggesting they become disordered upon activation and nucleosome binding. Nevertheless, the NTR may still have a functional role in the activated state as the AutoN motif has been suggested to support ISWI activity by interacting with the so-called suppressor helix (suppH) on Lobe1 in the nucleosome-bound state (15,16).

Comparing active and inactive state of the ATPase domain, it is clear that activation of ISWI requires a dramatic conformational change of the two ATPase lobes enabling the remodeler to bind ATP and the nucleosomal DNA. Typically, this conformational change is illustrated taking Lobe1 as the vantage point resulting in a $\sim 135^\circ$ swing of Lobe2 from a 'closed' into an 'open' state. We here would like to point out the benefits of taking Lobe2 as the vantage point, as illustrated in Figure 1B. From this perspective it becomes clear that the NTR blocks the DNA binding site of Lobe 2. Release of the NTR opens up this site to bind the nucleosomal DNA and releases the 'lock' to allow a $\sim 135^\circ$ rotation of Lobe1 such that the DNA and nucleotide binding sites are aligned on both lobes. Thus, the conformational change to the active state is not so much a *swinging out* of the lobes, but rather an *internal rotation*. Interestingly, this couples activation to remodeling, as also the remodeling reaction depends on the rotation of the lobes with respect to each other, which is thought to be driving the sliding of DNA. In addition, this view emphasizes the role of NTR release as a key step in activation and rationalizes the role of DNA binding herein.

Here, we address the conformational properties of the NTR in isolation, in the autoinhibited state of ISWI and upon binding to various activating epitopes using binding assays and NMR spectroscopy. We find that an extended ISWI ATPase domain including the NTR and NegC is in the autoinhibited state in solution, unable to bind an H4 tail peptide. The NTR is largely bound to the ATPase lobes as expected, but a long region that includes the AT-hook motif is highly dynamic in this state. By fluorescence anisotropy (FA) binding assays, we show that DNA binding induces a

conformational change that allows additional binding of H4. NMR data indicate that the AT-hook region is strongly affected by DNA binding and that H4 tail binding increases the dynamics of this altered conformation. In all states studied, large parts of the NTR are unobservable indicating they remain associated to the ATPase core, in line with the low stability of the isolated NTR domain. Overall our data identify the AT-hook region within the NTR as a highly dynamic element responsive to DNA and H4 tail binding. We propose that ISWI activation is induced by DNA binding to the NTR, loosening the interaction to Lobe2 and thus allowing for H4 tail binding.

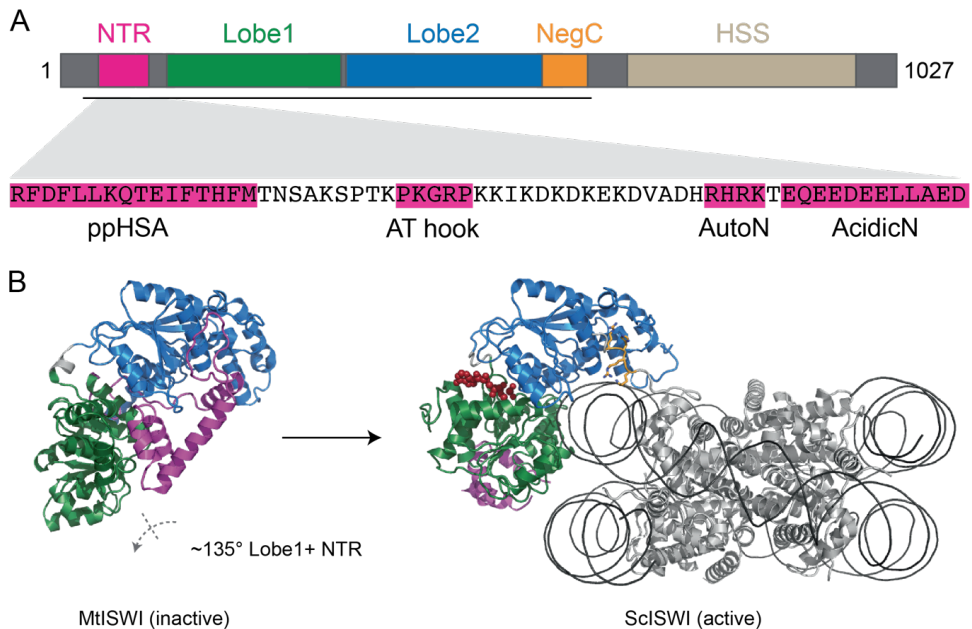


Figure 1. ISWI domains and structure. (A) Domain architecture of DmISWI. Conserved motifs in the N-terminal region (NTR) are highlighted in the sequence. The ISWI construct used in this study is indicated by the black line (residues 26-648). (B) Structure of MtlSWI (PDB entry: 5JXR, left) in closed conformation and nucleosome-bound structure of ScISWI (*Saccharomyces cerevisiae* ISWI, PDB entry: 6JYL, right) including ADP-BeFx (firebrick). Lobe2 has the same orientation in the two structures, Lobe1 and the NTR are rotated as indicated by the dashed arrow. Lobes 1 are colored green, Lobes 2 in blue, the NTR in magenta (only partly resolved in the active state) and the H4 tail in orange.

RESULTS

ISWI-ATPASE DOMAIN MUST BE DNA-BOUND TO ALLOW H4 TAIL BINDING

The order of ISWI activation by its nucleosomal epitopes was investigated by fluorescence anisotropy (FA) using an extended DmlSWI ATPase construct (residues 26-648, here simply referred to as ISWI), known to function as an autonomous remodeling machine (3). This construct contains the N-terminal region including the AutoN and AcidicN motifs, as well as the NegC region. Incubation of a fluorescein-labeled H4 tail peptide (residues 9-25) with ISWI did not result in significant changes in fluorescence anisotropy up to an ISWI concentration of $\sim 120 \mu\text{M}$, both in absence and presence of ADP-BeFx, a non-hydrolyzable ATP-mimic (Fig. 2A). This absence of binding starkly contrasts with the K_D of $13.5 \mu\text{M}$ previously reported for the interaction between a short H4-tail peptide (residues 13-21) and ISWI Lobe2 (11). The dramatic difference in binding affinity is most likely caused by use of the complete ATPase domain including the NTR in our experiment, indicating that the H4 tail binding site on Lobe2 is not or less available and that the H4 tail alone is unable to compete off the NTR. We conclude that the H4 tail or ADP-BeFx alone are not sufficient to change the conformation of ISWI into the active state.

Incubation of fluorescently labeled dsDNA (26 bp) with ISWI showed comparable affinities in presence and absence of the ATP-mimic (K_D 23 vs $16 \mu\text{M}$, respectively) (Fig. 2B), in line with previous reports (17). Notably, even if the Lobe2 DNA-binding surface is partly occluded by the NTR in the autoinhibited conformation, the Lobe1 DNA binding surface is exposed. In addition, the NTR itself features a patch of positively charge residues that is exposed in this state.

We next reasoned that DNA binding might induce a conformation change, as alluded to in Figure 1B, and thereby allow binding of the H4 tail peptide to ISWI. However, a direct binding assay of the ISWI-DNA complex and labeled H4 tail peptide would be complicated by DNA-binding of the H4 tail peptide. Therefore, we set up a competition assay, starting from a DNA-H4-tail complex, titrating in ISWI ATPase domain. The 26-bp dsDNA fragment binds the FAM-labeled H4 tail peptide with a K_D of $\sim 95 \mu\text{M}$ (Fig.

2C), so we chose this concentration as a starting point for the competition assay. In absence of ADP-BeFx, addition of ISWI ATPase causes a severe drop in the anisotropy signal due to precipitation (Fig. 2D). This means that a complex is formed involving the labeled H4 tail peptide, ISWI and the DNA, as we know that ISWI without DNA does not bind to the H4 tail and the DNA-H4 complex is soluble. Remarkably, addition of ADP-BeFx prevents this precipitation resulting in a further increase of the anisotropy signal. While small, this increase indicates formation of a stable complex including ISWI, DNA and H4 tail. We thus conclude that DNA binding causes a structural change in ISWI that permits the H4 tail to interact, requiring ADP-BeFx to stabilize the complex. Assuming that this state reflects an active conformation, this means that DNA precedes H4 tail binding in release of autoinhibition.

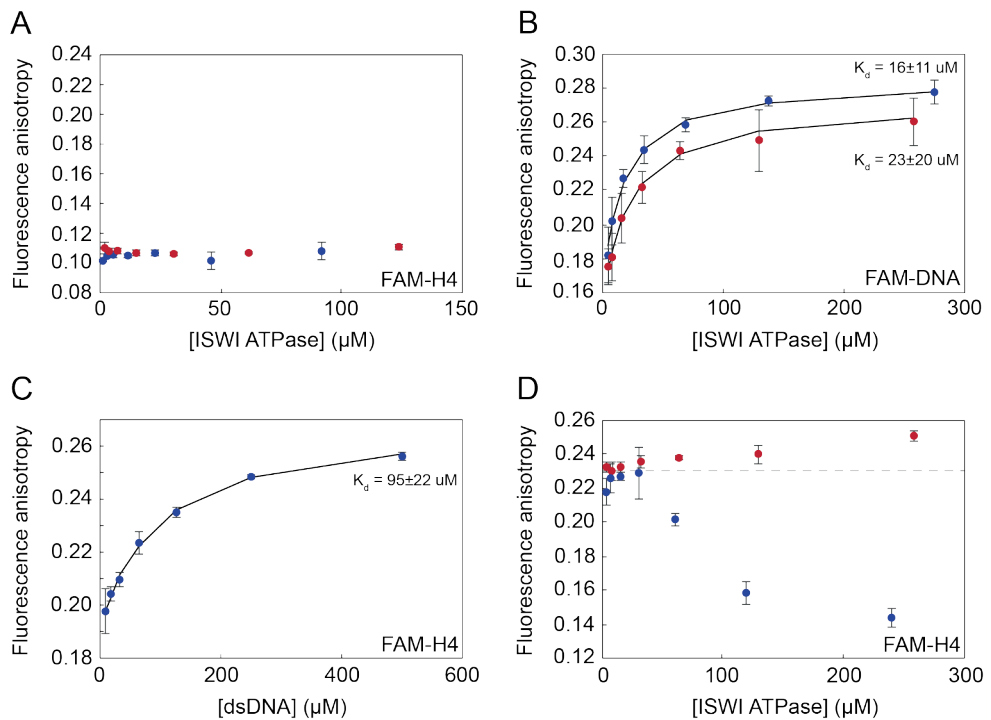


Figure 2. Fluorescence anisotropy binding assays of ISWI ATPase domain. (A) Titration of 10 nM N-terminally fluorescein-labeled H4 tail peptide with increasing concentrations of ISWI ATPase domain in presence (red) and absence (blue) of 1 mM ADP-BeFx. (B) Titration of 10 nM 5'- fluorescein-labeled dsDNA (26 bp) with increasing concentrations of ISWI ATPase domain in presence (red) and absence (blue) of 1 mM ADP-BeFx. The fits to determine the K_D are shown as black lines. Error estimation for the K_D is based on 1000 Monte-Carlo simulations. (C) Titration of 10 nM N-terminally fluorescein-labeled H4 tail peptide with

increasing concentrations of dsDNA (26 bp). The fit to determine the K_D is shown as a black line. Error estimation for the K_D is based on 1000 Monte-Carlo simulations. (D) Competition assay starting from 10 nM N-terminally fluorescein-labeled H4 tail peptide in presence of 90 μ M dsDNA and titrating with increasing concentration of ISWI ATPase domain in presence (red) and absence (blue) of 1 mM ADP-BeFx. The dotted grey guideline indicates the initial anisotropy. Error bars represent the standard deviation from three independent measurements.

THE NTR IS AN INTRINSICALLY UNSTABLE, MOSTLY DISORDERED DOMAIN

To understand what happens to the NTR at a structural level during these different stages of activation, we first characterized the structural properties of the NTR as an isolated domain, as the range of conformational changes it experiences may be reflected in its intrinsic structural properties. While the NTR is partly folded as an α -helical bundle in the autoinhibited state, this may largely be due to close interaction with the ATPase lobes. Indeed, secondary structure predictions for the NTR sequence by several different algorithms show the same trend of intrinsic disorder with some potentially α -helical stretches (see Fig. 4B). We thus designed a construct of the NTR (ISWI residues 23-110 and an additional GPAW N-terminal insertion to allow absorbance measurements) and studied this protein fragment to assess its intrinsic structural propensities.

Purification of the ISWI-NTR was not straightforward, as the protein tends to form soluble aggregates under a wide range of conditions, as observed in analytical gel filtration profiles, and tends to precipitate over time at high concentrations. Thermal shift assays revealed a clear increase of fluorescence with a transition midpoint of $\sim 55^\circ\text{C}$ in various phosphate or Tris buffers of pH 7 to 8, which could reflect melting of a stable tertiary structure (Fig. 3A). We thus concluded that there is little buffer dependence of NTR stability and continued characterization in gel-filtration buffer to avoid an additional buffer-exchange step. Circular dichroism (CD) measurements at 15, 25 and 35°C showed mostly a disordered profile, although data at wavelengths < 200 nm could not be reliably measured due to the buffer conditions (Fig. 3B). Of note, the distinctive maximum at 222 nm for helical proteins was not observed. The top 3 proteins with a similar CD profile (based on lowest area differences) are 47%

irregular/disordered, 34% sheet and 19% helical. The folding state of the NTR was also visualized in a 'double wavelength' plot ($[\Theta]_{222}$ versus $[\Theta]_{200}$) (Fig. 3C) (18,19). This shows that at all three temperatures measured the NTR is closest to a pre-molten globule state, exhibiting some amount of residual secondary structure.

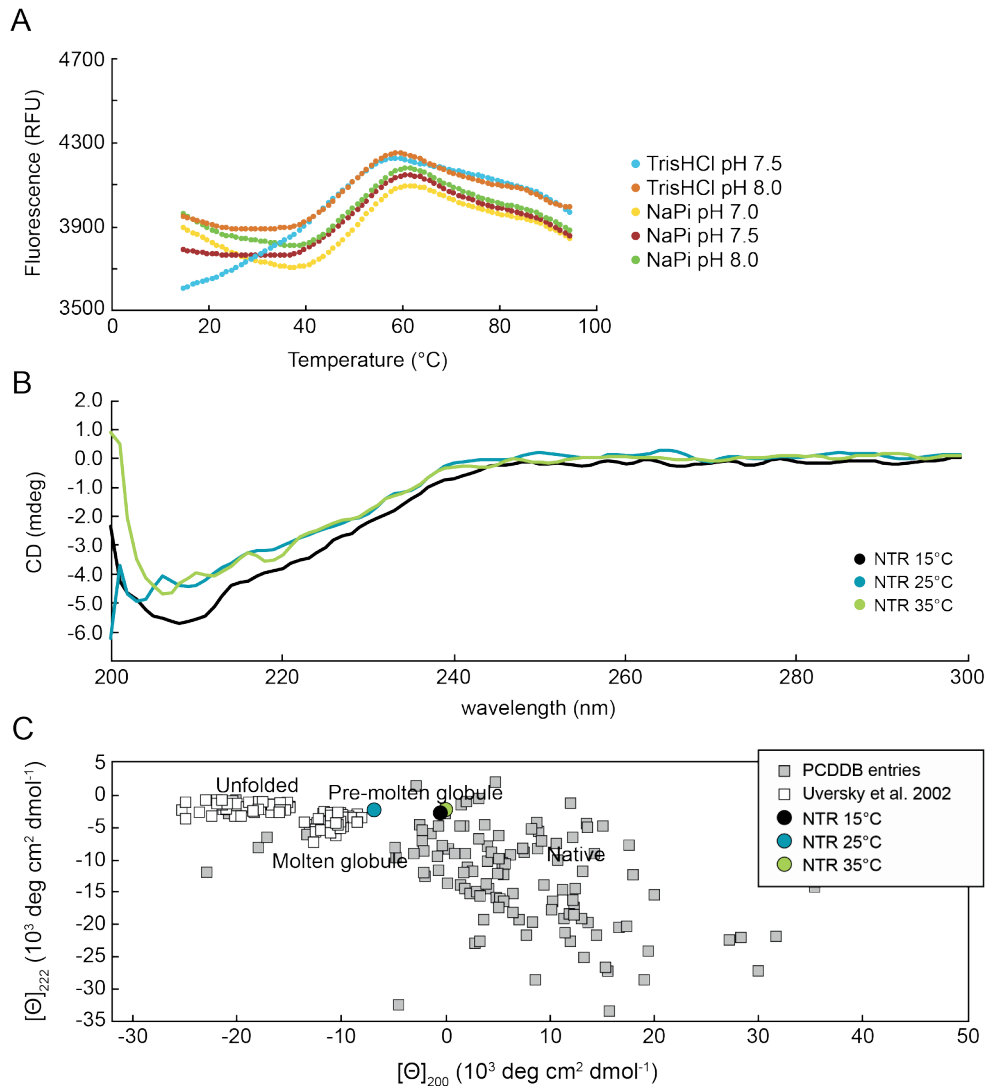


Figure 3. Thermal stability and secondary structure of ISWI NTR. (A) Blank corrected thermal shift assay data using SYPRO Orange in the indicated buffer conditions at 150 mM ionic strength. (B) Circular dichroism spectra at 15, 25 and 35°C, obtained from 4 accumulations and corrected for buffer absorbance at matching temperature. Data recorded in 50 mM Tris

pH 7.0, 150 mM KCl, 1 mM EDTA, 1 mM DTT. (C) Double wavelength plot of $[\Theta]_{222}$ versus $[\Theta]_{200}$ from CAPITO analysis. Color coding indicated in the figure.

To assess the structural properties of the NTR at a residue-specific level, we prepared ^{15}N and ^{13}C -labeled NTR and studied this by NMR. Across several buffer conditions tested, the NMR fingerprint spectrum showed a narrow proton chemical shift dispersion, extensive peak overlap and a heterogeneous pattern of peak intensities, indicative of a largely disordered protein. In phosphate buffer at pH 7, additional weak, dispersed signals could be observed, suggesting some regions of secondary structure (Fig. 4A). Spectra of different sample batches yielded slightly different spectra, showing peak doublings and minor chemical shift changes. In addition, soluble aggregates isolated by gel filtration showed a similar NMR peak pattern, but with extensive line broadening due their larger molecular weight. These observations suggest that the NTR free in solution is present in multiple conformational or aggregational states. Backbone chemical shift assignment was hampered by low signal intensities and missing correlations in the 3D assignment spectra. In total, 46% of the residues could be confidently assigned, corresponding approximately to the first 20 and last 30 residues of the construct. Notably, no assignments could be obtained for the central 30 residues, which are folded into α -helices in the MtlSWI structure (see Fig. 4B). For the assigned parts, the difference between the $\text{C}\alpha$ and $\text{C}\beta$ secondary chemical shifts was used as a measure of secondary structure. The close correspondence to the random coil shifts strongly suggests that the assigned residues are largely disordered, including large parts of the AcidicN region that is folded as a helix in the MtlSWI structure (Fig. 4B). For a number of residues, including K94 of the AutoN region, multiple resonances could be assigned in the HSQC spectrum, indicating a heterogeneous conformational state of the NTR.

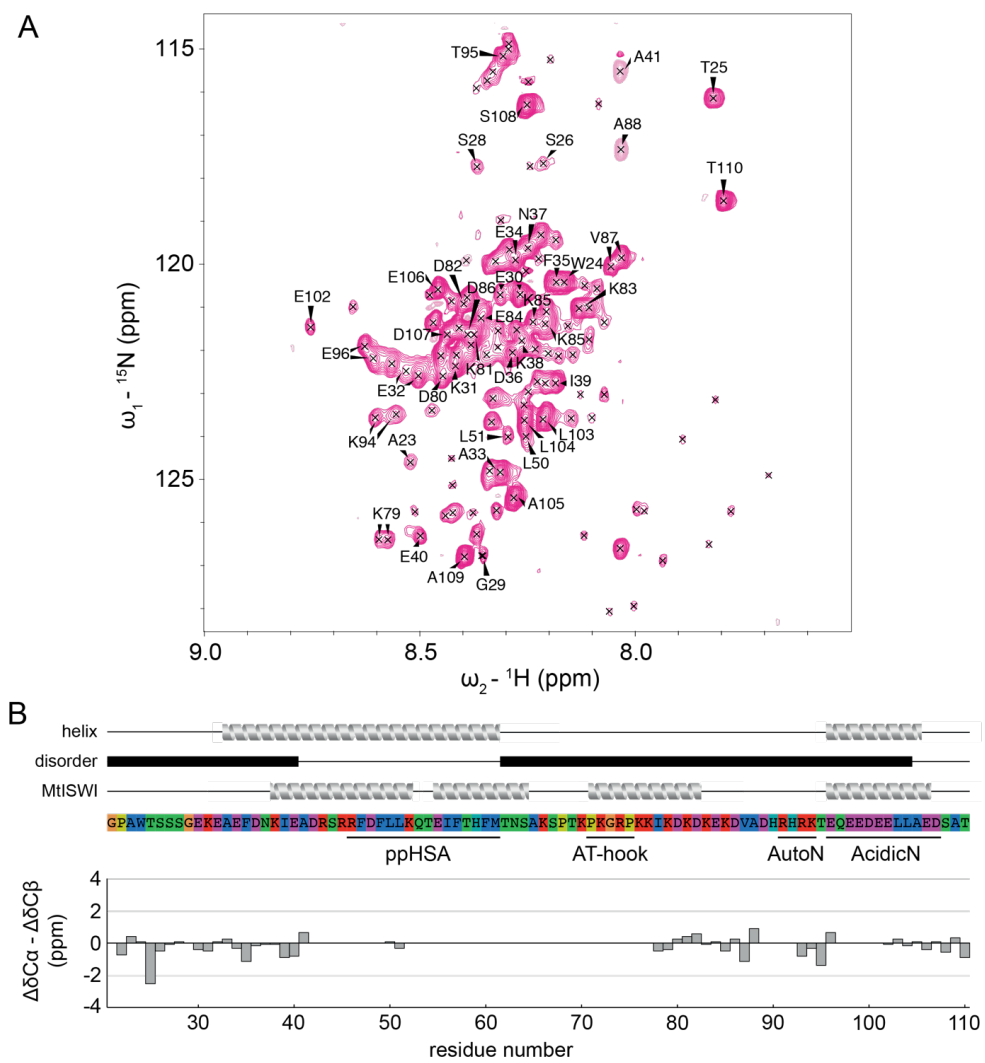


Figure 4. NMR fingerprint spectrum and secondary structure analysis of ISWI NTR. (A) 2D ^{15}N - ^1H HSQC, backbone region with assignments, recorded in 50 mM NaPi, pH 7.0, 150 mM NaCl, 1 mM PMSF, 1 mM DTT and 5% D_2O . (B) Predicted secondary structure (grey helices, top) and disordered regions (black boxes) in the ISWI NTR compared to the helices found in the inactive state structure of MtISWI (PDB entry: 5JXR). The lower panel shows a bar graph of the difference between the secondary $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts of the assigned NTR residues. Amino acids in the sequence are color-coded by type: hydrophobic residues in blue, basic residues in red, acidic residues in pink, hydrophilic residues in green, histidines in teal, prolines in yellow and glycines in orange. Conserved sequence motifs are indicated underneath the sequence.

Overall, our data show that the isolated NTR is a mostly disordered domain that is unstable in solution and prone to aggregate over a wide range of conditions. We suspect that the central part may be (transiently) folded in helical form and induce aggregation into larger molecular weight oligomers, whose melting may have been observed in the TSA assays. Clearly, the α -helical fold as observed in the available structures of the NTR is not observed in the isolated domain, so the context of the ATPase domain is required to stabilize this fold. The unstable nature of the NTR also means that activation of ISWI is unlikely to result in full release of the NTR as a flexible tail, as this might destabilize the protein. We thus conclude that the NTR will likely maintain (transient) interactions to the ISWI core or nucleosome surface.

LIMITED DYNAMICS OF THE NTR IN THE ISWI ATPASE DOMAIN

We next used the full ISWI ATPase construct including the NTR and NegC autoinhibitory domains in an NMR study to assess the conformational state of the domain and in particular the NTR. Using a fractionally deuterated sample and TROSY-based experiments to boost sensitivity, a backbone NH fingerprint could be recorded revealing many and well-dispersed resonances, indicating the protein is well-folded (Fig. S1). We here focused on the dynamic parts of the protein as we aimed to determine whether the NTR is released from the core upon activation. Solution NMR experiments are most sensitive to dynamic regions, in particular for large proteins such as the 75 kDa ISWI ATPase domain. To emphasize the dynamic editing, we made use of protonated proteins in which rapid ^1H relaxation further suppresses the signals of the folded and rigid parts. The resulting spectrum is dominated by intense resonances in the center of the spectrum, whose intensity and close to random coil chemical shift indicate that these belong to highly dynamic parts of ISWI. Using the deuterated protein sample, assignments for 48 of these resonances could be obtained (by Vincenzo Lobbia) (Fig. 5A). Assigned resonances correspond mostly to a stretch of residues in the very N-terminus of the construct leading up to the $\alpha 1$ helix (K31-K38) and a region comprising the AT-hook and leading up to the AutoN motif (K66-A88). With the exception of residues 71-82, which would be expected to form the $\alpha 3$ helix, the observation that these residues are dynamic agrees with the inactive MtlSWI structure. Notably, since the sequence of DmlSWI contains two prolines in the $\alpha 3$ region, this helix is not likely to be conserved in our construct. Analysis of the backbone chemical

shifts indeed indicates that all observed regions are in random coil conformation (Fig. 5D). In addition to the NTR, the assigned resonances correspond to a stretch of residues at the very C-terminus of the protein, thus covering the NegC region, as well as residues in the first loop in Lobe2, indicating that these regions are highly dynamic within the inhibited protein (Fig. 5C). As only a small number of the intense resonances in Figure 5A remained unassigned, a major part of the NTR is motionally restricted, as it is folded and/or bound to the ISWI core.

Of the 21 NTR residues that were assigned in both ATPase domain and isolated NTR, 8 NTR residues have similar chemical shifts in the two samples, while the remaining 13 residues, mostly those close to the folded parts of NTR in MtlISWI, have significantly different chemical shifts (Fig. 5B and E). A number of residues assigned to the NTR, mostly in the region of the AT-hook (residues 66-77), could not be assigned in the free NTR due to its conformational heterogeneity, but give rise to clear and intense cross peaks within the context of the ATPase domain. This suggests that the complete AT-hook region between the $\alpha 2$ and AcidicN helices is stabilized in a disordered loop conformation within the context of the ISWI protein. Vice versa, residues in the AutoN and AcidicN motifs that could be assigned in the free NTR, are not assigned in the ATPase domain, likely due to their folding in the full protein. Folding into α -helices is assumed for DmlISWI based on sequence homology with MtlISWI. Overall, our data indicate that the free ISWI protein is in the autoinhibited state in solution and that only the AT-hook region in DmlISWI NTR forms a highly dynamic loop in this state.

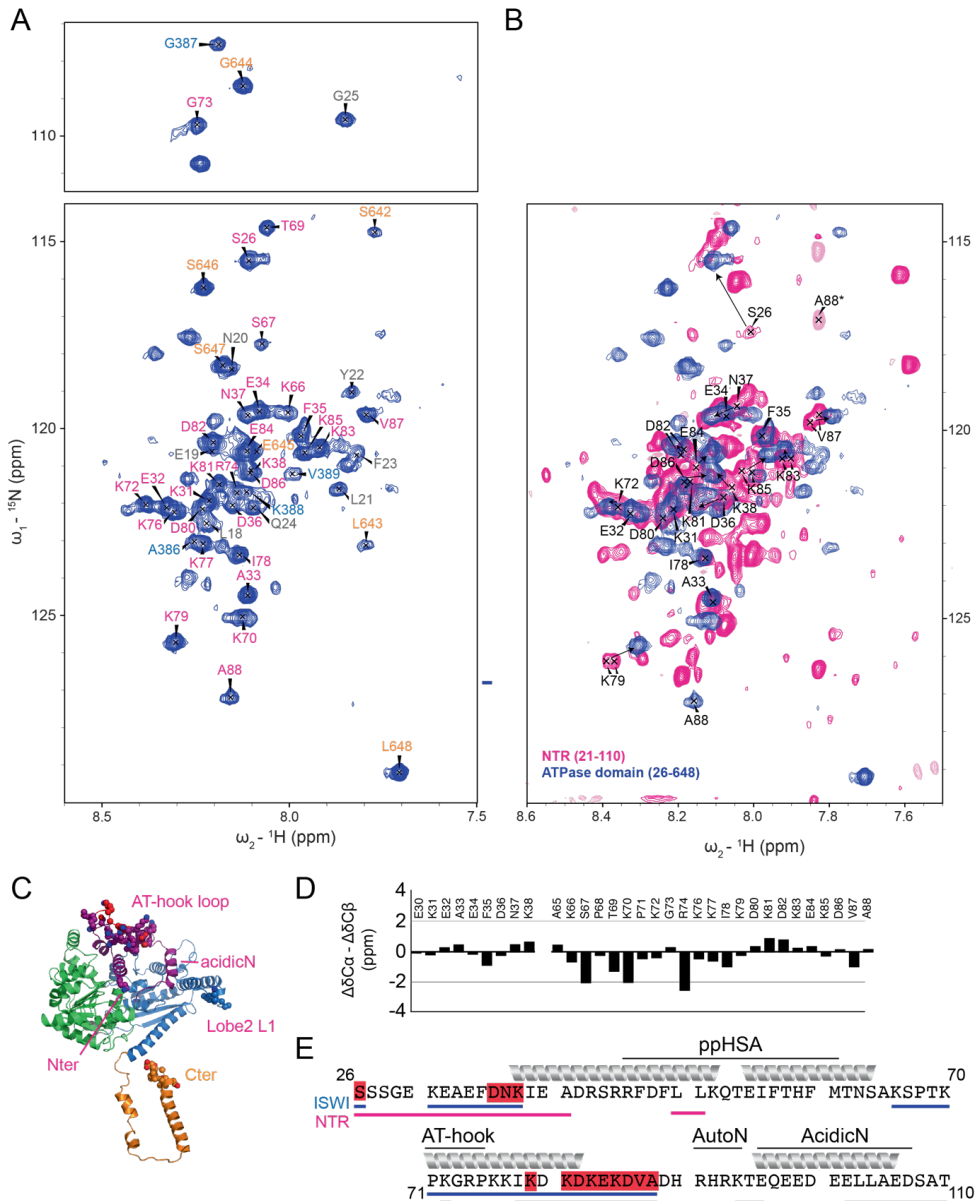


Figure 5. NMR on the ISWI ATPase domain. (A) 2D NH TROSY spectrum of ^1H , ^{15}N -labeled ISWI ATPase domain with assignments indicated. Assignment labels are color-coded according to the domain they belong to (matching Fig. 1 and panel B here); grey labels refer to the N-terminal TEV cleavage site. (B) Overlay of the HSQC spectra of ISWI ATPase domain (blue) and to ISWI NTR (pink). Residues with an assignment available for both spectra are labelled. (C) Location of assigned ISWI residues (shown in spheres) indicated on a homology model of DmISWI based on the MtISWI structure. Color-coding as in Fig. 1. Note that the AT-

hook region is modelled as a loop, consistent with our NMR data. (D) Difference between the secondary $C\alpha$ and $C\beta$ chemical shifts for NTR within the extended ATPase construct. Values close to zero indicate a random coil conformation. (E) Summary of the changes in the NTR region due to ATPase domain context. Residues assigned in free NTR and in the extended ATPase construct are underlined in magenta and blue, respectively. Residues that show chemical shift perturbation (see panel B) are shaded in red. Secondary structure of the NTR in the MtISWI structure is displayed above the sequence.

DNA BINDING CAUSES A STRUCTURAL REARRANGEMENT THAT IS NOT UNDONE BY H4 TAIL BINDING

To obtain a more detailed understanding of the conformational changes during ISWI activation, we performed a series of NMR titration studies. Since ADP-BeFx is thought to be important to reach the activated state and we found that ADP-BeFx stabilizes the DNA-H4-ISWI complex, we performed most interaction studies in the presence of the ATP mimic. Addition of ADP-BeFx alone to the ATPase domain did not significantly change peak intensities, and only some minor chemical shift changes were observed, ruling out a major conformational change in the flexible (visible) parts of ISWI, such as release of the NTR (Fig. S2). Titrations of H4 tail peptide to the ISWI ATPase domain in presence or absence of ADP-BeFx showed no changes in chemical shift or intensity of the observed resonances (Fig. S3), in agreement with the lack of binding observed in FA. Based on the FA studies, we inferred that DNA binding causes a conformational change that allows H4 tail binding. To test whether this change causes direct release of the NTR from the H4 tail binding site on Lobe2 into solution or whether the NTR remains associated to the core ATPase domain, we titrated first DNA and then H4 tail peptide to ISWI in our NMR setup that probes dynamic regions only.

Addition of a 26-bp dsDNA fragment up to one molar equivalent compared to ISWI caused major changes in the overall peak pattern and resulted in significantly fewer intense resonances compared to the unbound state (Fig. 6A). Even though the protein will not be fully saturated at this DNA concentration (190 μ M), the titration was stopped to limit the signal loss due to precipitation as well as the contribution of non-specific DNA binding. Most resonances that show clear chemical shift perturbations are exchanging in the slow regime (see insets Fig. 6A). Several of these were tracked to their new position based on slight chemical shift changes, correspondence in line shapes and overall peak pattern. Analysis of the peak intensities and chemical shift

changes in unbound and bound states shows that residues in the very N-terminal region of the NTR have strongly reduced peak intensity but no chemical shift changes, while residues in the AT-hook region have both reduced intensities and altered chemical shifts (Fig. 6B). Notably, a number of residues in the AT-hook region, including K66, K72 and K77, disappear from the spectrum. Since several new resonances are observed in the DNA-bound spectrum, this likely means that these residues experience larger changes in their chemical environment but could not be reliably tracked to their new position. Together, these data suggest that the AT-hook region either directly binds DNA or that it experiences a conformational change as a result of DNA binding elsewhere. Given the basic character of this region and the well-known DNA binding of the AT-hook motif (20), direct binding seems most likely. Analysis of the chemical shift perturbations suggests that the off-rate k_{off} is less than $\sim 50 \text{ s}^{-1}$, assuming 1:1 binding (see insets in Fig. 6A). Under the assumption that the observed affinity in the FA experiments ($\sim 20 \text{ }\mu\text{M } K_D$) corresponds to binding of the AT-hook, one would derive an on-rate of $< \sim 2 \cdot 10^6 \text{ s}^{-1} \text{ M}^{-1}$, which seems rather slow for an electrostatic interaction. Nevertheless, since the FA experiment measures an overall, apparent affinity, it may well be that the AT-hook-DNA interaction is actually more affine.

In addition to the clear effects of DNA-binding in the NTR, also the dynamic regions in Lobe2 and NegC are affected. Resonances for residues in the Lobe2 L1 loop and the C-terminus are strongly reduced in peak intensity upon DNA binding, while those of the C-terminal residues, see L648 in Figure 6A, also experience chemical shift changes and peak splitting (Fig. 6B). Since these regions are unlikely to be directly involved in DNA binding because of their sequence, these observations point to an indirect effect of DNA binding. As these regions are distant from each other and from the NTR, this argues for global conformational change upon DNA interaction.

Subsequent addition of H4 tail peptide up to 10 molar equivalents caused only minor CSPs, which is in particular notable for the NTR AT-hook region residues (Fig. 6C). This indicates that the H4 tail peptide is not able to compete off the DNA and thereby undo the changes observed in the DNA titration. Given the excess of peptide and the K_D determined by FA, the protein should be close to saturated with H4. Indirect evidence for H4 tail binding comes from analysis of the peak intensities. Remarkably, resonances for the AT-hook region become more intense upon addition of H4 tail, while other residues do not show a significant change (Fig. 6B). Assuming the H4 tail binds

to the Lobe2 site, this intensity increase could result from release of the AcidicN helix from Lobe2 and thus increased flexibility of the attached AT-hook region. Since no new resonances are observed, the AutoN and AcidicN regions must be bound to another site on the ATPase core. In addition, the C-terminal resonance shows a gradual transition from a split peak, of which one at the unbound position, to only a peak at the shifted position. This suggests that H4 tail binding stabilizes the alternate conformation that is already induced by DNA binding.

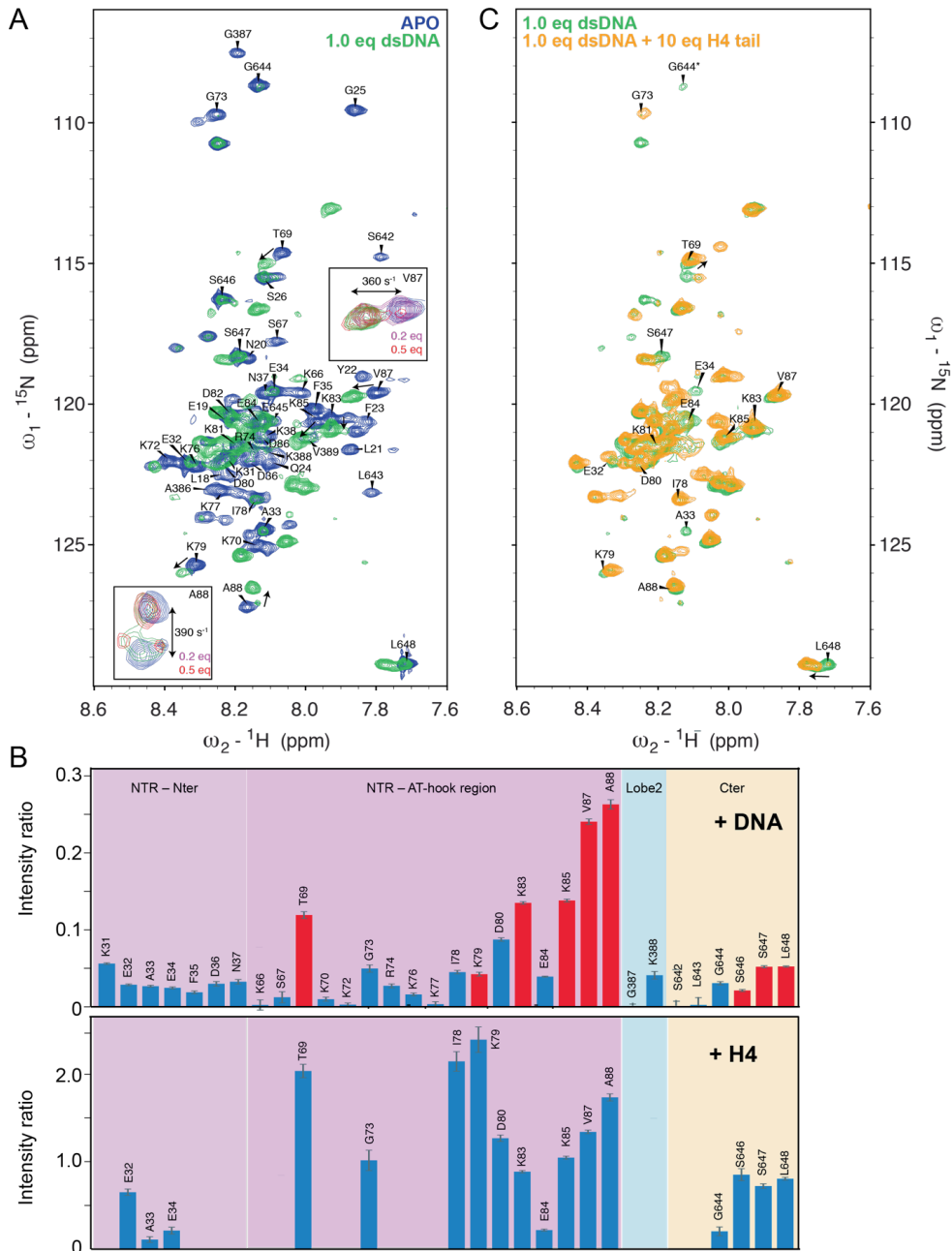


Figure 6. NMR on the ISWI ATPase domain. (A) 2D NH HSQC spectrum of ISWI ATPase domain in presence of 1 mM ADP-BeFx (blue) and after addition of 1.0 equivalent of dsDNA (26bp) in presence of ADP-BeFx (green). Assignments in the blue spectrum are indicated, CSPs are indicated by arrows. Insets show titration overlay for two resonances in slow exchange. (B) Bar plot of intensity ratios for confidently assigned peaks upon DNA binding

(top) and additional H4 tail binding (bottom). Red colored bars in the top panel indicate residues with significant CSPs. Overall peak intensity ratios are low due to precipitation, especially in the first two titration steps. (C) 2D NH HSQC spectrum of ISWI ATPase domain in presence of 1 mM ADP-BeFx after addition of 1.0 equivalents of dsDNA (26bp) (green) and after additional titration with 10 equivalents of H4 tail peptide (orange). Confident assignments for non-overlapping peaks are indicated.

DISCUSSION

In this chapter we identified the AT-hook region in the NTR as a highly dynamic and responsive element in ISWI activation based on a series of FA binding assays and NMR studies. The transition from the autoinhibited to activated state upon nucleosome binding is a crucial step to allow remodeling. This transition hinges on a large conformational change of the NTR to align the ATPase lobes and allow for H4 tail binding. We showed that NTR domain itself is intrinsically unstable and largely disordered. Within the context of the ISWI protein, only the very N-terminus and the AT-hook region of the NTR are disordered and highly dynamic. Our data indicate that the AT-hook region can bind DNA and that DNA binding induces a conformational change that allows subsequent H4 tail binding. Tail binding does not result in release of the NTR as an extended floppy tail, but the assumed helical regions of the NTR rather remain associated to the ATPase core, in line with the low intrinsic stability of the NTR domain.

Even if the NMR experiments were tuned to probe only the highly dynamic regions of ISWI, they nevertheless provide valuable support for the integrative model of DmISWI in the autoinhibited state (11). In particular, our solution data support the NegC domain being collapsed against the ATPase core and highlight that the AT-hook region, that is folded as an α -helix in MtlISWI, forms a flexible loop. Since this loop is connecting the first two NTR helices that are bound to Lobe1, to the AutoN and AcidicN regions that are bound to Lobe2, that raises the question whether this loop will also allow for some breathing of the lobes. Perhaps DmISWI is not as rigidly locked in the autoinhibited state as the thermophilic MtlISWI. Furthermore, we found that the L1 loop in Lobe2 is highly dynamic. Interestingly, this loop was not resolved in the nucleosome-bound structure of ISWI by Yan *et al.* (15), suggesting that this loop has intrinsic flexibility in both autoinhibited and activated state.

While the Lobe2 DNA binding site is blocked by the NTR, examination of a model for DmISWI in the autoinhibited state highlights several basic patches that could bind DNA, including the Lobe1 nucleosomal DNA binding site, a basic patch on Lobe2 and the AT-hook region in NTR. As we could only observe the AT-hook region in the DNA titration NMR experiment, we cannot exclude that the DNA in fact bound to other sites than to the AT-hook. However, as the chemical shift changes were close to saturation at one molar equivalent of DNA, the majority of the DNA should at least be close to this region. An indirect effect of DNA-binding on the AT-hook conformation cannot be excluded, e.g. the observed changes could result from DNA competing off AcidicN from the DNA binding site on Lobe2, followed by a rearrangement of AcidicN and the AT-hook loop on the ATPase surface. Alternatively, the AT-hook loop directly binds DNA. Notably, the DNA binding site could also be extended to the adjacent and basic AutoN motif. This could be a functional interaction to orient the inactive remodeler on the nucleosome and loosen the AutoN/AcidicN–Lobe2 interaction, to ultimately allow invasion of the H4 tail.

Similarly, the NMR data provided indirect evidence for binding of the H4 tail to Lobe2 through the observed increase in dynamics of the AT-hook loop that is proximal to the tail binding site. Notably, a few studies have suggested that the H4 tail can directly interact with the HSS domain (12,21), but this domain is not present in our construct. In the FA study we found that

ATP-BeFx binding enhances the stability of the ISWI-DNA-H4 complex in solution. Nucleotide binding however does not seem to be required for activation, as recently a ‘primed’ state of ISWI bound to the nucleosome in absence of nucleotides was captured by cryo-EM (14).

Since ISWI is bound to a separate and short piece of DNA and H4 tail rather than an intact nucleosome, it allowed to separate their effects. Yet, it raises the question whether these separate fragments are also able to induce the global rearrangement of the ATPase lobes from the inactive to the active state. Some support for a global conformational change comes from gradual chemical shift changes observed for the C-terminal residues upon binding of DNA *and* H4 tail. Only upon binding of the two compounds the unbound conformation is fully converted to the “bound” conformation. The peak intensity decreases for the L1 loop in Lobe2 could also reflect a global rearrangement, but as this loop is near the Lobe2 DNA binding site, these changes could result from the DNA binding itself.

In conclusion, we propose that ISWI activation is a three-step process (Fig. 7). The remodeler first binds to DNA, which causes large changes in the flexible parts of the protein and partial rearrangement of the lobes. Autoinhibition by the NTR is 'loosened' by DNA binding to the AT-hook loop and basic AutoN motif. Once ISWI binds the nucleosome in proximity of the H4 tail, this tail can bind to Lobe2 to stabilize the interaction and promote further release of the NTR and complete rearrangement of the ATPase lobes. The NTR then moves to bind a different site on the ATPase surface, for example the SuppH motif to support remodeling activity by ISWI, as suggested previously (15), or alternatively to a site on the nucleosome, i.e. the acidic patch, as was also suggested (22,23).

A key aspect in this model is that NTR release by competition with the H4 tail is only possible after weakening the NTR-Lobe2 interaction. While still to be confirmed in future work, this model and our work place the NTR at the center of the release of autoinhibition and highlight the interplay between multiple dynamic interactions in achieving activation.

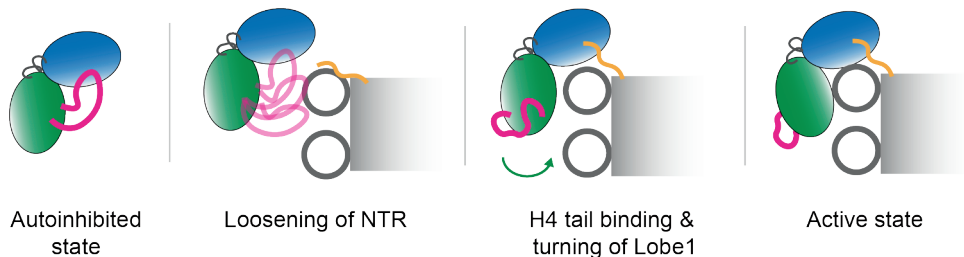


Figure 7. Schematic view of ISWI activation by binding of DNA and H4 tail. Lobe1 is shown in green, Lobe2 in blue, NTR in pink, H4 tail peptide in orange and the nucleosome in grey.

MATERIALS & METHODS

EXPRESSION AND PURIFICATION OF ISWI ATPASE DOMAIN

The ATPase domain of *D. melanogaster* ISWI (residues 26-648) was expressed from a pProEX-Htb vector including an N-terminal His6-TEV tag (plasmid constructed in Christoph Müller lab, EMBL Heidelberg) in *E. coli* BL21(DE3) pLysS cells in either LB medium or M9 medium supplemented with $^{15}\text{NH}_4\text{Cl}$, glucose, vitamins and trace

elements. After reaching an OD₆₀₀ of 0.6-0.8, the main cultures were cooled down at 4°C for 15 minutes. Expression was induced by addition of 0.2 mM IPTG and allowed to continue overnight at 18 °C. Cells were harvested by centrifugation, resuspended in ice-cold T300 buffer (50 mM Tris, pH 7.4, 300 mM NaCl). After addition of 20 mM imidazole, 0.5 mM BME and protease inhibitor cocktail, the cells were lysed by sonication on ice (6x 10 s, 10 s off in between, 10 μm amplitude) and the lysate was centrifuged at 20000 rpm for 30 min at 4°C. ISWI ATPase domain was purified on a 5-mL HisTrap HP column (GE Healthcare) in combination with a 5-mL HiTrap Q column (GE Healthcare) in 15 mM TrisHCl pH 7.4, 150 mM KOAc, 1 mM BME and 20-400 mM imidazole. The pooled peak fractions were dialyzed to buffer without imidazole in a 12-14 kDa MWCO dialysis membrane. The retrieved pool was concentrated and loaded onto a HiLoad Superdex 200 16/600 pg gel filtration column equilibrated in 20 mM Hepes KOH, pH 7.6, 200 mM KCl, 0.2 mM EDTA, 1 mM BME. Peak fractions from gel filtration were pooled, concentrated and buffer exchanged to NMR buffer (20 mM KPi, pH 6.8, 100 mM KCl, 10 mM DTT). All purification and dialysis steps were performed at 4°C.

EXPRESSION AND PURIFICATION OF ISWI NTR

The N-terminal region of *D. melanogaster* ISWI (residues 23-110) was expressed from a pGeX-6P1 vector including an N-terminal GST-tag (plasmid constructed by Johanna Ludwigsen, LMU Munich) in *E. coli* BL21(DE3) cells in either LB medium or M9 minimal medium supplemented with ¹⁵NH₄Cl and/or ¹³C-glucose. Expression was induced by addition of 0.5 mM IPTG at OD₆₀₀=0.6 and allowed to continue for 4h at 37°C. Cells were harvested by centrifugation, resuspended in ice-cold lysis buffer (50 mM Tris, pH 8.0, 300 mM KCl, 20% sucrose, 0.2% TritonX-100, 0.5 mM PMSF, 1 mM DTT, protease inhibitor cocktail) and lysed using lysozyme and French press. The lysate was centrifuged at 18000 rpm for 25 min at 4°C and the resulting supernatant was loaded onto a 5-mL GSTrap HP column (GE Healthcare) equilibrated in lysis buffer. The column was washed in 5 column volumes (CV) of wash buffer (50 mM Tris, pH 8.0, 300 mM KCl, 20% sucrose, 0.5 mM PMSF, 1 mM DTT) followed by 10 CV of cleavage buffer (50 mM Tris pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT). PreScission protease (GE Healthcare) was loaded onto the column (200 U in 5 mL of cleavage buffer) and the column was incubated overnight at 4°C. Cleaved ISWI-NTR

was collected the next day by washing the column with 3CV of cleavage buffer. The entire purification procedure of ISWI-NTR was performed at 4°C. ¹³C¹⁵N-labeled ISWI-NTR for 3D NMR experiments was dialyzed to NMR buffer (50 mM NaPi, pH 7.0, 150 mM NaCl) in a 3.5 kDa MWCO dialysis membrane.

FLUORESCENCE ANISOTROPY

All binding reactions contained 10 nM of either FAM-labeled double-stranded DNA (FAM-TGTTAATACCCCGGGCTTTATTAC) or FAM-labeled H4 tail peptide (5-FAM-KGLGKGGAKRHRKVLRLD-NH₂) and the indicated amount of unlabeled binding partner in MES buffer (20 mM MES, pH 6.8, 100 mM KCl, 10 mM DTT) in a total volume of 10 μL. For the competition assay, all samples contained 10 nM FAM-labeled H4 tail peptide, 90 μM unlabeled dsDNA and the indicated amount of ISWI ATPase domain in MES buffer. All samples were prepared in triplicate and contained 0.1 mg/mL BSA to reduce sticking of the protein and peptide to the plastic tips and wall of the plate. Fluorescence anisotropy measurements were carried out at an excitation wavelength of 485 nm and emission wavelength of 535 nm at room temperature on a SpectraMax i3 plate reader (Molecular Devices) using a black 384-well plate.

THERMAL SHIFT ASSAY

Thermal shift assay was performed in a transparent 96-well plate with each well containing 2.5 μM ISWI-NTR, 5x SYPRO Orange, 50 mM buffer and 150 mM ionic strength. Six different buffer conditions were included by using two buffer components (TrisHCl and NaPi at 50 mM) and three pH values (7.0, 7.5 and 8.0). Total ionic strength was adjusted for ionic strength of the buffer component and kept constant at 150 mM. All samples were prepared in duplicate and a 'no protein' and 'no dye' control was included in every run. Thermal shift assay was run by putting the sealed plate in a RT-qPCR machine in FRET mode, using a temperature range of 15-95°C at 1°C /min.

CIRCULAR DICHROISM

UV/CD spectra were recorded on a JASCO J-810 circular dichroism spectropolarimeter using a 0.1 cm path length QS cuvette (Hellma Analytics). ISWI-NTR samples were prepared at 15 μM concentration in cleavage buffer (50 mM Tris

pH 7.0, 150 mM KCl, 1 mM EDTA, 1 mM DTT). Spectra were recorded from 200 to 300 nm, with a data pitch of 1 nm and scanning speed of 50 nm/min. Spectra were averaged over four accumulations and the data were analyzed using CAPITO (18). Sample was measured at 15, 25 and 35°C with a waiting time of 15 minutes in between, to allow the sample to adjust to the new temperature. Buffer samples were recorded at matching temperature and subtracted from the measured sample curve.

NMR SPECTROSCOPY

NMR experiments were carried out on Bruker Avance III spectrometer at 900 MHz (ISWI ATPase experiments and ISWI NTR assignment) or Bruker Avance III-HD spectrometer at 850 MHz (other ISWI NTR spectra) at 288 K (NTR) or 293 K (ATPase domain) unless indicated otherwise. The sample for assignment of ISWI-NTR contained ~200 μ M of $^{13}\text{C}^{15}\text{N}$ -labeled protein in 50 mM NaPi, pH 7.0, 150 mM NaCl, 1 mM PMSF, 1 mM DTT and 5% D_2O . Assignments were made based on HNC0, HNCA, HNCACB and CBCACONH spectra recorded at 288 K. Assignment was 46% complete (42 out of 90 observed resonances). NMR titration experiments on ISWI ATPase domain were performed using ~190 μ M protonated ^{15}N -labeled protein in presence of 1 mM ADP-BeFx, a 1 mM dsDNA (26 bp) stock and a 5.0 mM H4 tail peptide stock, all in MES buffer (20 mM MES, pH 6.8, 100 mM KCl, 10 mM DTT) and 5-10% D_2O . Chemical shift perturbations and intensity changes were monitored by recording subsequent ^{15}N - ^1H 2D TROSY spectra.

All spectra were processed in NMRPipe or Bruker Topspin and analyzed in Sparky. Prediction of ISWI-NTR backbone torsion angles and secondary structure propensity based on chemical shifts were calculated using the TALOS+ server (24).

SECONDARY STRUCTURE AND DISORDER PREDICTIONS

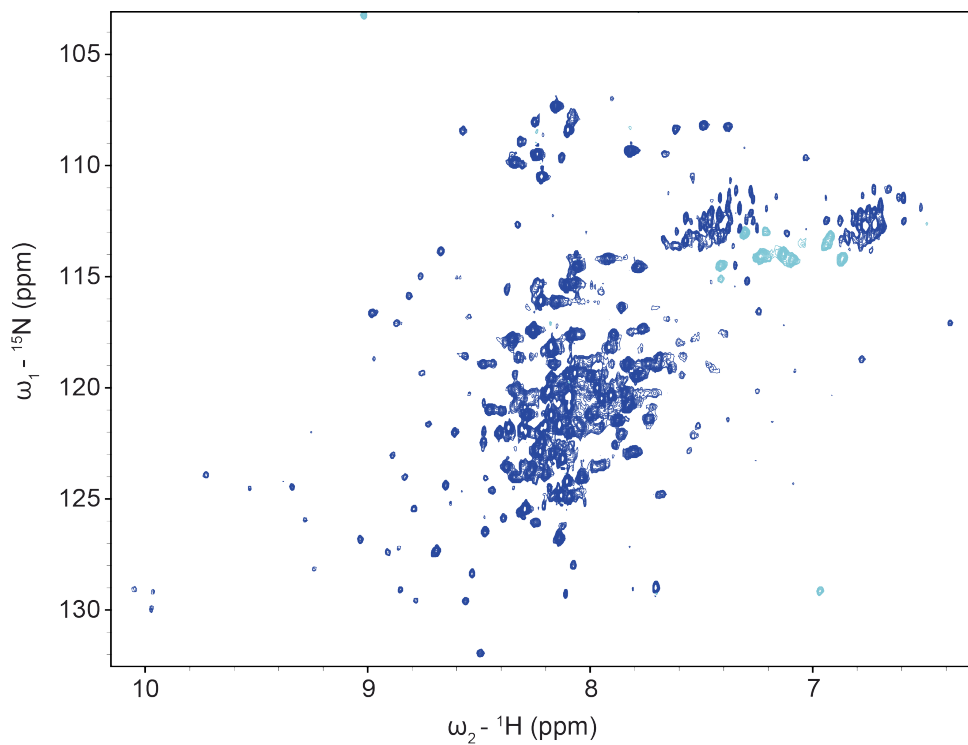
Secondary structure and disorder predictions for ISWI NTR were performed using Phyre2 (25), SCRATCH protein predictor (<http://scratch.proteomics.ics.uci.edu>), IUPred (26), JPred (27), PONDR (<http://www.pondr.com>) and the CFSSP server (28).

REFERENCES

1. Flaus, A., Martin, D.M., Barton, G.J. and Owen-Hughes, T. (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res*, **34**, 2887-2905.
2. Clapier, C.R. and Cairns, B.R. (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem*, **78**, 273-304.
3. Mueller-Planitz, F., Klinker, H., Ludwigsen, J. and Becker, P.B. (2013) The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. *Nat Struct Mol Biol*, **20**, 82-89.
4. Grune, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F., Becker, P.B. and Muller, C.W. (2003) Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell*, **12**, 449-460.
5. Yamada, K., Frouws, T.D., Angst, B., Fitzgerald, D.J., DeLuca, C., Schimmele, K., Sargent, D.F. and Richmond, T.J. (2011) Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature*, **472**, 448-453.
6. Leonard, J.D. and Narlikar, G.J. (2015) A nucleotide-driven switch regulates flanking DNA length sensing by a dimeric chromatin remodeler. *Mol Cell*, **57**, 850-859.
7. Clapier, C.R. and Cairns, B.R. (2012) Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature*, **492**, 280-284.
8. Ludwigsen, J., Pfennig, S., Singh, A.K., Schindler, C., Harrer, N., Forne, I., Zacharias, M. and Mueller-Planitz, F. (2017) Concerted regulation of ISWI by an autoinhibitory domain and the H4 N-terminal tail. *Elife*, **6**.
9. Dang, W., Kagalwala, M.N. and Bartholomew, B. (2006) Regulation of ISW2 by concerted action of histone H4 tail and extranucleosomal DNA. *Mol Cell Biol*, **26**, 7388-7396.
10. Racki, L.R., Naber, N., Pate, E., Leonard, J.D., Cooke, R. and Narlikar, G.J. (2014) The histone H4 tail regulates the conformation of the ATP-binding pocket in the SNF2h chromatin remodeling enzyme. *J Mol Biol*, **426**, 2034-2044.
11. Yan, L., Wang, L., Tian, Y., Xia, X. and Chen, Z. (2016) Structure and regulation of the chromatin remodeller ISWI. *Nature*, **540**, 466-469.
12. Harrer, N., Schindler, C.E.M., Bruetzel, L.K., Forne, I., Ludwigsen, J., Imhof, A., Zacharias, M., Lipfert, J. and Mueller-Planitz, F. (2018) Structural Architecture of the Nucleosome Remodeler ISWI Determined from Cross-Linking, Mass Spectrometry, SAXS, and Modeling. *Structure*, **26**, 282-294 e286.
13. Armache, J.P., Gamarra, N., Johnson, S.L., Leonard, J.D., Wu, S., Narlikar, G.J. and Cheng, Y. (2019) Cryo-EM structures of remodeler-nucleosome intermediates suggest allosteric control through the nucleosome. *Elife*, **8**.
14. Chittori, S., Hong, J., Bai, Y. and Subramaniam, S. (2019) Structure of the primed state of the ATPase domain of chromatin remodeling factor ISWI bound to the nucleosome. *Nucleic Acids Res*.

15. Yan, L., Wu, H., Li, X., Gao, N. and Chen, Z. (2019) Structures of the ISWI-nucleosome complex reveal a conserved mechanism of chromatin remodeling. *Nat Struct Mol Biol*, **26**, 258-266.
16. Liu, X., Li, M., Xia, X., Li, X. and Chen, Z. (2017) Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature*, **544**, 440-445.
17. Chin, J., Langst, G., Becker, P.B. and Widom, J. (2004) Fluorescence anisotropy assays for analysis of ISWI-DNA and ISWI-nucleosome interactions. *Methods Enzymol*, **376**, 3-16.
18. Wiedemann, C., Bellstedt, P. and Gorlach, M. (2013) CAPITO--a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics*, **29**, 1750-1757.
19. Uversky, V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, **11**, 739-756.
20. Aravind, L. and Landsman, D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Research*, **26**, 4413-4421.
21. Hwang, W.L., Deindl, S., Harada, B.T. and Zhuang, X. (2014) Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature*, **512**, 213-217.
22. Gamarra, N., Johnson, S.L., Trnka, M.J., Burlingame, A.L. and Narlikar, G.J. (2018) The nucleosomal acidic patch relieves auto-inhibition by the ISWI remodeler SNF2h. *Elife*, **7**.
23. Dann, G.P., Liszczak, G.P., Bagert, J.D., Muller, M.M., Nguyen, U.T.T., Wojcik, F., Brown, Z.Z., Bos, J., Panchenko, T., Pihl, R. *et al.* (2017) ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature*, **548**, 607-611.
24. Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*, **44**, 213-223.
25. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, **10**, 845-858.
26. Dosztanyi, Z., Csizmek, V., Tompa, P. and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433-3434.
27. Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*, **43**, W389-394.
28. Kumar, T.A. (2013) CFSSP: Chou and Fasman Secondary Structure Prediction server. *Wide Spectrum*, **1**, 15-19.

SUPPLEMENTARY INFORMATION



4

Figure S1. 2D NH TROSY spectrum of fractionally deuterated ${}^{15}\text{N}$ -labeled ISWI ATPase domain recorded at 900 MHz.

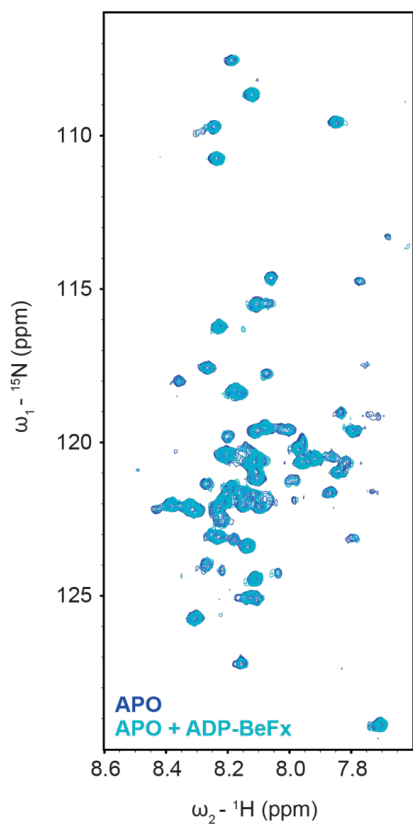


Figure S2. 2D NH HSQC spectrum of ISWI ATPase domain in absence (blue) and in presence of 1 mM ADP-BeFx (light blue).

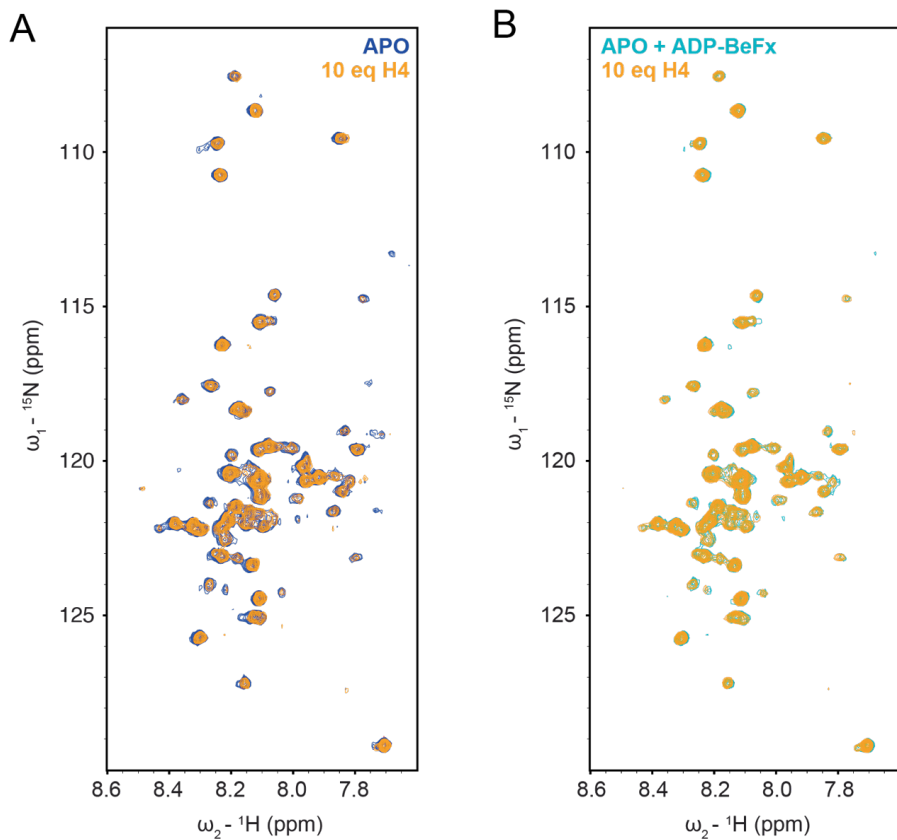


Figure S3. NMR titration overlays of ISWI ATPase domain with H4 tail peptide. (A) 2D NH HSQC spectrum of ISWI ATPase domain (blue) and after addition of 10 equivalents of H4 tail peptide (orange). (B) 2D NH HSQC spectrum of ISWI ATPase domain (light blue) and after addition of 10 equivalents of H4 tail peptide (orange).

SEGMENTAL ONE-STRAND ISOTOPE-
LABELING OF NUCLEOSOMAL DNA-601 AS
A TOOL IN NMR STUDIES OF
NUCLEOSOME INTERACTIONS

ABSTRACT

The nucleosomal DNA plays a crucial role in nucleosome interactions with chromatin remodelers and pioneer transcription factors as well as epigenetic reader, writer and eraser domains. NMR studies of nucleosome-protein interactions have proven to be able to provide novel insights in the molecular basis of these interactions, and their structural and dynamical impact. Yet, the nucleosomal DNA has so far remained invisible in high-resolution NMR studies due to limitations in specific isotope-labeling methods. Here, we developed a method for segmental one-strand isotope labeling of six thymine residues in the nucleosomal DNA DNA-601 sequence. The limited number of labeled residues reduces signal overlap and the labeling of thymines, which contain a methyl group, creates the opportunity of using MeTROSY NMR experiments to improve signal to noise ratios in large complexes like the nucleosome. We describe the synthesis of the segmentally isotope-labeled DNA601 in NMR sample quantities as well as assignment of the six methyl signals. After incorporation of the labeled DNA into nucleosomes, we performed a titration with aclarubicin, an anti-cancer drug that intercalates into the DNA minor grooves, showing that the labeled thymines can serve as local probes to observe site-specific binding events in the context of the nucleosome. This opens up new opportunities to study the contribution of nucleosomal DNA to nucleosomal interactions in molecular detail.

INTRODUCTION

The nucleosomal DNA plays a crucial role in many nucleosome-protein interactions. These interactions involve many proteins such as chromatin remodelers and pioneer transcription factors that have DNA as their primary target site (1,2). Next to these, epigenetic reader, writer and eraser domains that primarily target the histones can also bind the close-by DNA to enhance their affinity and specificity (3-6). DNA-binding chromatin factors may cause changes in DNA conformation, histone-DNA contacts and nucleosome flexibility ('breathing') that are essential for the biological effect of the interaction. This is especially clear for ATP-dependent nucleosome remodelers that need to weaken and break histone-DNA contacts to be able to move the nucleosome along the DNA (2,7). Even if the molecular mechanism of remodeling is still under debate, it is very likely to involve extensive deformation of the nucleosomal DNA in the form of bulging or loop propagation. To obtain a better understanding of the molecular mechanism underlying these and other types of DNA-dependent processes, we aimed to study these using NMR spectroscopy, exploiting its unique sensitivity for dynamic and weak binding events.

In general, to perform NMR experiments on nucleosomes, specific isotope-labeling schemes are used to be able to observe the histones inside the complex and monitor changes upon binding of an interacting protein (see Chapter 1). This strategy has been successful in determining binding interfaces between two proteins as well as interfaces between a protein and the DNA, although only observed from the protein side. For direct observation of the changes in the nucleosomal DNA upon binding of a protein, the DNA needs to be isotope-labeled. Uniform isotope-labeling can be achieved by producing the DNA from a multi-repeat plasmid in *E. coli* in minimal medium with ^{13}C -glucose and $^{15}\text{NH}_4\text{Cl}$ as the sole carbon and nitrogen sources, respectively, or by PCR in presence of isotope-labeled nucleotides (8). However, as double-stranded DNA is composed of only four building blocks and has a very even secondary structure, the resulting limited chemical shift dispersion will lead to extensive signal overlap. In addition, the rigidity of the DNA and its incorporation into the 200 kDa nucleosome will cause major line broadening and very poor signal-to-noise due to relaxation.

To overcome both of these hurdles, we adopted a strategy for site-specific isotope-labeling of nucleosomal DNA. We here report the specific labeling of six thymine residues at superhelix location (SHL) +6 of the nucleosomal DNA-601 sequence (Fig. 1). Labeling only six residues in one strand will drastically reduce signal overlap. In addition, thymines are the only nucleotides that contain a methyl group, which allows the use of MeTROSSY NMR experiments to enhance signal-to-noise for high molecular weight systems like the nucleosome (9). SHL+6 was picked as the labeling site, as this is close to the binding region of nucleosome remodeler ISWI and several other nucleosome-interacting proteins around SHL-2, on the neighboring DNA gyre (3,10,11). The six thymines will thus function as local probes to report on binding and dynamics in this area.

In this chapter, we describe the synthesis of the segmentally isotope-labeled DNA601 in NMR sample quantities, followed by assignment of the methyl group signals in a short fragment of DNA-601 as well as the full-length nucleosomal sequence. Subsequently, this full-length labeled DNA-601 was incorporated into nucleosomes and the feasibility of NMR studies on this system explored. To demonstrate that the labeled thymines can be used as probes for local structure perturbation of the DNA, we performed a titration of the nucleosomes with aclarubicin, an anti-cancer drug that intercalates into the DNA minor groove (12).

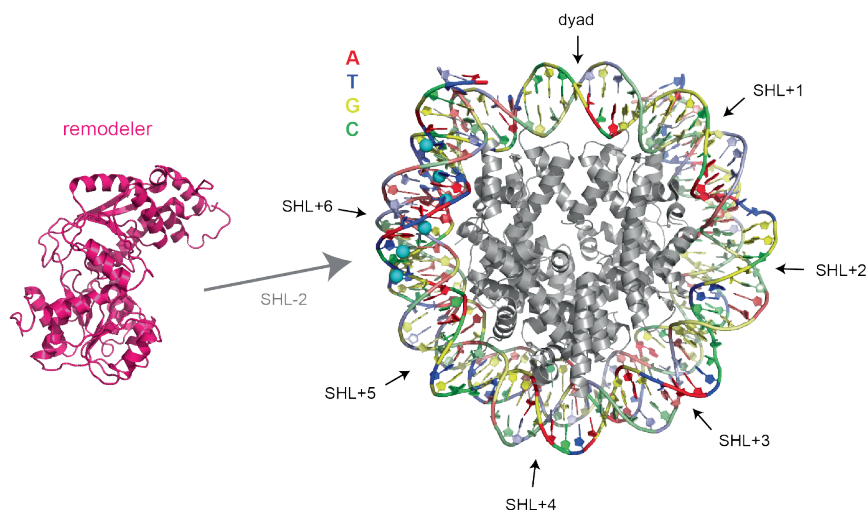


Figure 1. Position of the labeled thymines in the nucleosomal DNA. Methyl groups of the labeled thymines are shown in cyan spheres. Nucleotides are color-coded according to base

type: thymidine in blue, adenosine in red, guanosine in yellow and cytosine in green. Histone octamer is shown in grey, an interacting remodeler in pink.

RESULTS

To obtain segmentally isotope-labeled DNA-601 for NMR experiments, we needed to adopt a labeling strategy that allows large-scale synthesis of the DNA in mg quantities. Three methods for large-scale production were described in Chapter 3 of this thesis: plasmid-based production from a multi-repeat plasmid, large-scale PCR and ramified RCA. As mentioned above, plasmid-based production only allows for uniform isotope-labeling by supplementing the culturing medium with isotope-labeled nutrients. Ramified RCA could be performed with one of the primers containing the desired isotope-labels, however, as most of the repeats in the rRCA product are not initiated by a primer but by continuation from the previous monomer, this would lead to only a fraction of the product molecules containing the isotope-labeled segment. Therefore, we chose to use large-scale PCR for the synthesis of segmentally labeled DNA-601, using a 30-nt forward primer with six isotope-labeled thymines and an unlabeled reverse primer to achieve one-strand segmental isotope-labeling in the final product (Fig. 2).

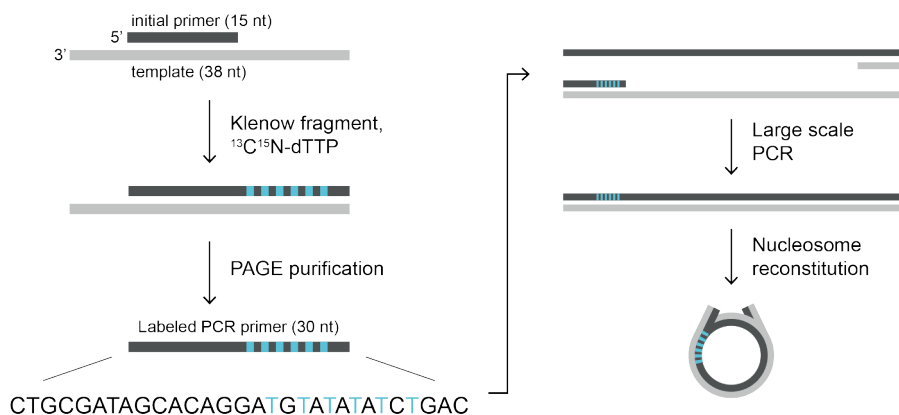


Figure 2. Schematic overview of the synthesis of segmentally isotope-labeled DNA601 and its incorporation into nucleosomes. An initial primer-template partial duplex is subjected to a fill-in reaction with Klenow fragment in presence of uniformly $^{13}\text{C}^{15}\text{N}$ -labeled dTTP and unlabeled dGTP, dCTP and dATP. The resulting 30 nt labeled PCR primer is PAGE purified and used in a large-scale PCR reaction to produce full-length nucleosomal DNA with the six labeled

thymines incorporated. The full-length segmentally isotope-labeled DNA is used to reconstitute nucleosomes.

LABELED PRIMER SYNTHESIS

The isotope-labeled primer was synthesized by performing a fill-in reaction in the presence of uniformly ^{15}N - and ^{13}C -labeled dTTPs and unlabeled dGTPs, dCTPs and dATPs using an initial template-primer partial duplex with the template having both a 3' and a 5' overhang (Fig. 2). During the reaction the 5' overhang is filled-in by the polymerase, while the 3' overhang remains. This ensures a large enough difference in size of the template and the labeled primer (38 vs. 30 nt) to facilitate separation on and purification from a denaturing PAGE gel. The lengths of the initial primer and template determine the 5' and 3' boundaries, respectively, of the segment that will be isotope-labeled. Therefore, varying these lengths allows tuning of which and how many thymines will be labeled. However, care must be taken to maintain the size difference for the template, initial primer and labeled primer to be able to separate them on gel, as this is crucial for purification. In addition, the final length of the labeled primer should not be too long, as this could compromise efficient annealing of primer-template duplex during the final PCR production of the nucleosomal DNA.

We chose to use the DNA-601 strong-positioning sequence with a total length of 167 bp, including 10 bp of linker DNA on each side. To achieve thymine isotope-labeling at SHL+6, we used a 38 nt initial template in combination with a 15 nt initial primer, that after fill-in yielded a 30 nt product, containing six isotope-labeled thymine residues (Fig. 3). The use of the Klenow (Large) fragment for polymerization ensures removal of possible non-templated adenine residues at the 3'-end of the labeled primer that could cause decreased priming efficiency during PCR. The yield was optimized by varying template-primer ratio (Fig. 3). A 1.25-fold excess of initial primer was chosen to limit the amount of left-over primer after the reaction. After gel purification and IEX chromatography, the yield was found to be 6-7 nmol of pure labeled primer per mL of fill-in reaction.

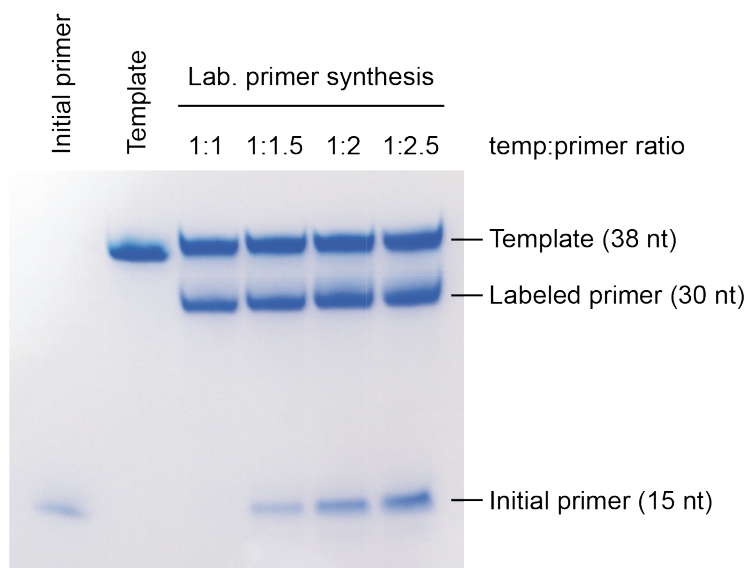


Figure 3. Denaturing PAGE gel of labeled primer synthesis. Ratios of template to initial primer concentration at the start of the reaction are indicated above each lane. The oligonucleotides present at the end of the reaction are well separated for purification of the isotope-labeled primer.

LARGE SCALE PRODUCTION OF 6T-LABELED DNA-601

The 30-nt segmentally isotope-labeled primer was used in large-scale PCR to obtain the full-length DNA-601 for nucleosome reconstitution. For the preparation of one nucleosome NMR sample, typically 2-3 mg of DNA is required. Normally, 40 96-well plates of 50 μ L PCR reactions run on at least 4 parallel thermocyclers are needed to reach this yield by PCR within 2-3 days (~200 mL reaction volume). To reduce the workload and the number of machines required, we used a 60-well thermo block on a Biometra machine that can accommodate 0.5 mL PCR tubes. To ensure proper temperature adjustment over the whole sample, the amount of time per PCR step was slightly increased and the tubes were filled up to 200 μ L only. This way, 12 mL of PCR reactions could be performed per run, with one full run taking approximately 2h. We were able to increase the yield to about 0.4 mg per 12 mL run, which means that it was feasible to produce 2 mg of DNA in 1-2 days on one thermocycler. The full-length DNA601 was purified from the reaction using IEX chromatography for use in NMR experiments.

ASSIGNMENT OF THE NMR METHYL SIGNALS

In order to use the isotope-labeled thymines as local probes in nucleosome-protein interactions, it is essential to assign the signals from its methyl groups. Assignment of short oligonucleotides is usually achieved by a so-called sequential walk in a ^1H - ^1H NOESY spectrum, exploiting the through-space coupling between the aromatic H6/H8 protons and the H1' proton on the sugar ring from the same residue, as well as the previous residue. However, due to extensive signal overlap and relaxation, this is not feasible for the 100 kDa full-length DNA-601. Alternatively, we used the partial duplex (21 kDa) before purification of the labeled primer as an intermediate to obtain the initial spectra. As the segment containing the labeled thymines is double-stranded in this fragment, we assumed that the chemical environment of the methyl groups would be highly similar to the full-length DNA-601.

In a $^1\text{H}^{13}\text{C}$ -HMQC spectrum, we indeed observed a cluster of peaks in the methyl region that originate from the thymine methyl groups in the partial duplex (see Fig. 6A). The proton chemical shift dispersion was limited, so the expected six individual peaks were not fully resolved. Recording a ^1H - ^1H NOESY yielded a spectrum that was still too crowded for assignment. To reduce the number of peaks in the NOESY spectrum, we recorded a ^{13}C -edited NOESY, filtering out all protons that are not connected to a carbon-13 in the direct dimension and therefore selecting only thymine resonances in the labeled segment (Fig. 4). This excludes the possibility of assignment via H6/H8-H1' connectivities between residues, however, we used chemical shift predictions in an attempt to identify the methyl signals. Chemical shifts were predicted using DSHIFT (13-15) and predicted peaks were categorized as low, medium and high intensity based on proton-proton distances of 0-3, 3-4 and 4-5 Å in a regular B-DNA structure of the sequence. The two methyl resonances in the GpT and CpT step could be assigned using their unique chemical shift signature in the H7-H6/H8 region of the ^{13}C -filtered NOESY spectrum (Fig. 4). The other four methyls occur in ApT steps; the spectral overlap does not permit their assignment based on chemical shift predictions.

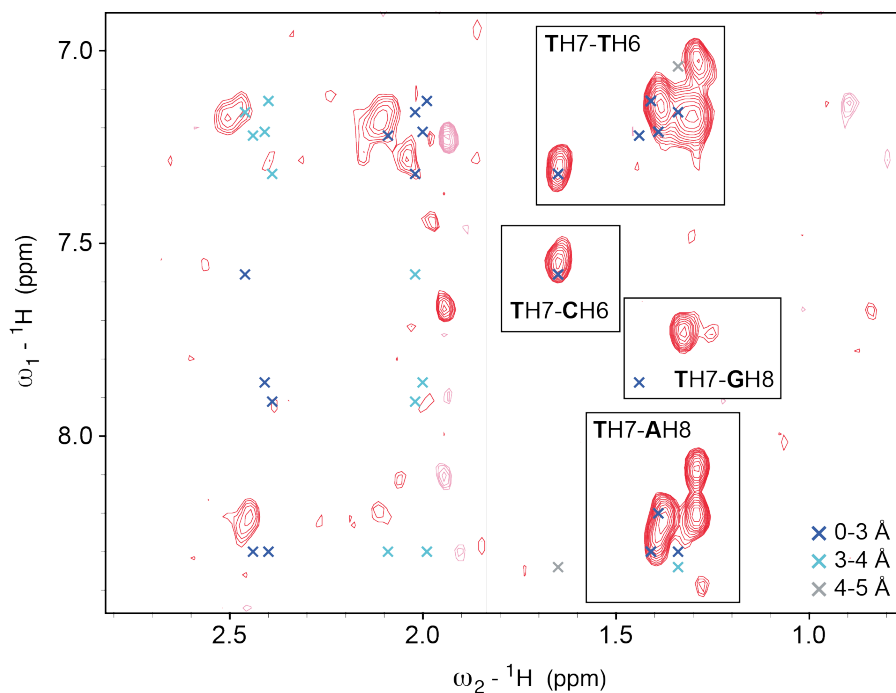


Figure 4. ^{13}C -filtered NOESY spectrum of the partial duplex (38-30nt) containing the six labeled thymine residues. The displayed region of the spectrum shows correlations between methyl signals and aromatic resonances. Predictions for cross-peaks are categorized according to their expected intensity (based on the distance between atoms in regular B-DNA). Cross-correlations for the labeled thymine methyl groups are indicated by black boxes.

In order to assign the remaining four thymine methyl resonances, we turned to a 15 bp unlabeled duplex of the same sequence to construct a ^1H - ^1H sequential walk using TOCSY and NOESY spectra. We assumed that assignments obtained this way would be transferrable to the spectrum of the labeled partial duplex and eventually to the full-length DNA-601. In the TOCSY spectrum, only through-bond coupled proton-proton resonances are observable. In the H6/H8-H1' region of the spectrum, this only yields peaks for intra-residue correlations between H5 and H6 protons from cytosine residues. The obtained cytosine H6 chemical shifts served as a starting point for the sequential walk, further aided by the fact that one of the cytosines (C30) is at the 3'-end of one strand and could therefore be identified directly. Once the sequential walks for each strand were completed, the assigned H6/H8 chemical shifts were used to

identify the thymine methyl resonances via the cross peaks between T-H7 and its own H6, as well as that of the preceding residue, confirming the assignments from the H6/H8-H1' walk (Fig. 5). The methyl group assignments could successfully be transferred to the labeled partial duplex spectrum (Fig. 6A, Fig. S1), completing the assignment.

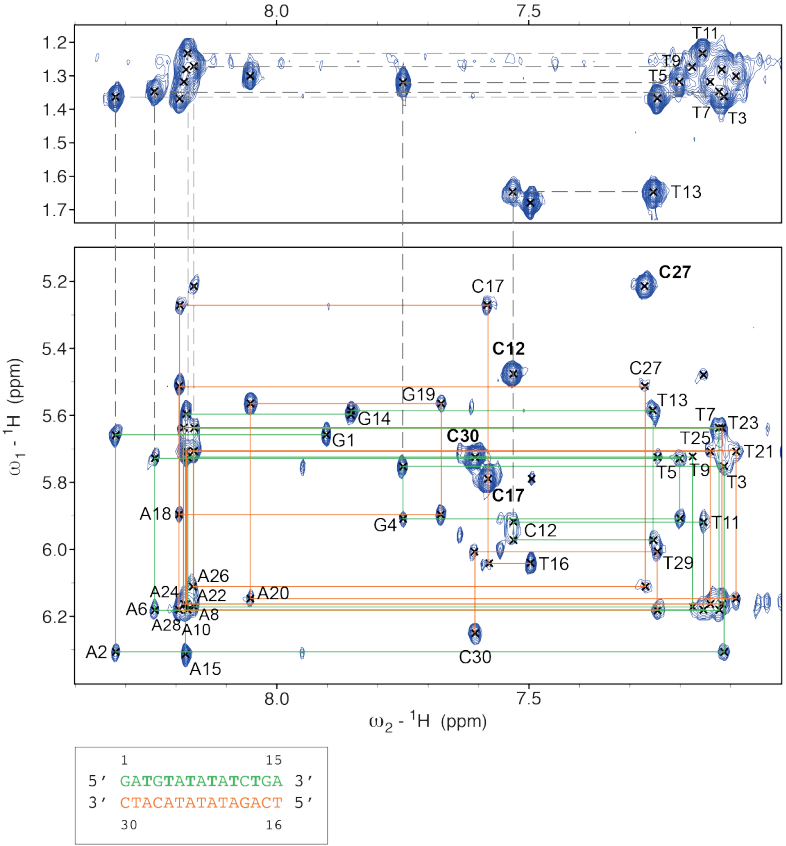


Figure 5. Assignment of a 15bp unlabeled duplex by sequential walk in $^1\text{H}^1\text{H}$ NOESY. The bottom panel shows the H1'-H6/H8 region with the sequential walks for the top strand (green, residue 1-15) and the bottom strand (orange, residues 16-30). H5-H6 cross peaks for cytosine are labeled in bold. The top panel shows the H7(Me)-H6/8 region of the same spectrum. Grey dotted lines show NOESY correlations of H6/H8 protons of thymine-preceding residues and the methyl resonances for the thymine residues that will be isotope-labeled.

MeTROSY spectra of the full-length segmentally labeled DNA-601 show a peak pattern very similar to the partial duplex, although the peaks are broader due to the higher molecular weight (Fig. 6B). To improve tumbling of the molecule and thereby reduce the line broadening, an additional spectrum was recorded at a higher temperature (35°C). Interestingly, this spectrum showed an additional peak or peak splitting around 1.4 ppm. This could be caused by alternative, sequence-dependent conformations of the (C)T residue that are in slow exchange on the NMR time scale. To optimize resolution further, we resorted to a constant-time HMQC (CT-HMQC) thereby removing broadening due to the $^1J_{CC}$ coupling, and recorded this at 45 °C. This yielded spectra with ~6 resolved peaks, that could now all be tentatively assigned based on the sequential walk chemical shifts, due the improved resolution (Fig. 6C).

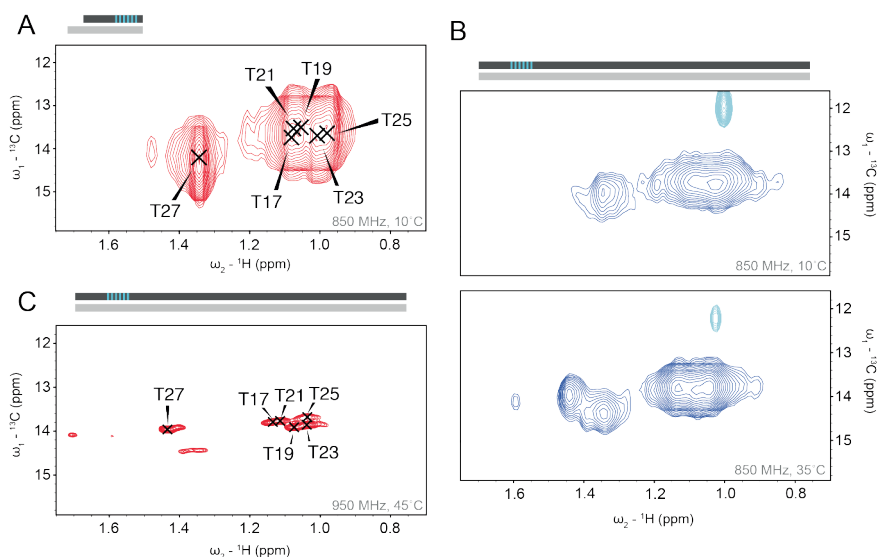


Figure 6. Methyl peak assignment of isotope-labeled thymines in the partial duplex and full-length DNA-601. (A) Methyl region of the HMQC spectrum of the partial duplex (38 + 30 nt) including assignments transferred from the sequential walk (see Fig. 5). (B) Methyl region of the HMQC spectrum of full-length DNA-601 (167 bp) at 10°C (top) and at 35°C (bottom). (C) Methyl region of the constant-time HMQC of full-length DNA-601 (167 bp) at 45°C, including transferred assignments. Thymine numbering according to the full-length sequence.

INCORPORATION OF SEGMENTALLY LABELED DNA601 INTO NUCLEOSOMES

The segmentally labeled DNA601 was used to reconstitute nucleosomes with *X. laevis* histone octamers. To minimize relaxation through the proton network of the histones, the octamers were fully deuterated and MeTROSSY spectra were recorded at 45°C (Fig. 7A). Bearing in mind that uniformly labeled thymidine was used in the preparation, it is notable that the only peaks that can be observed from the entire complex are the thymine methyl peaks, highlighting the benefit of using the methyl groups as probes. Interestingly, the additional peak near T27 that was observed for the free DNA-601 mostly disappeared, which suggests that the DNA conformation is stabilized inside the nucleosome.

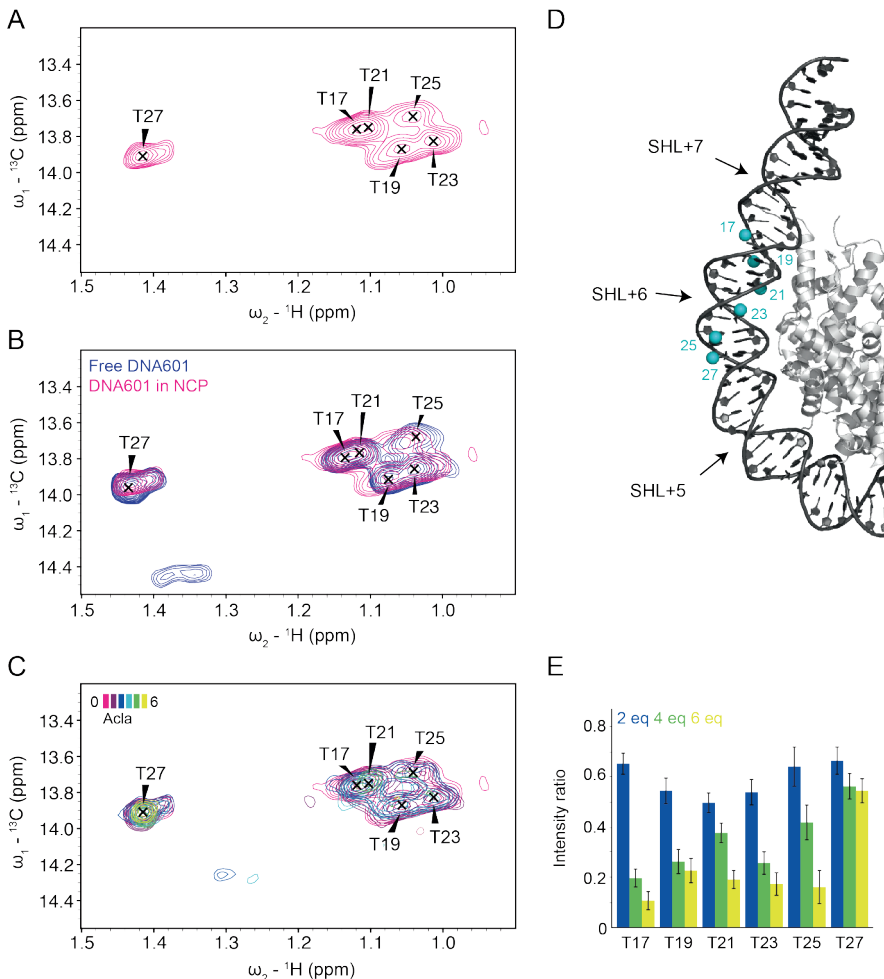


Figure 7. Titration of the nucleosome with aclarubicin. (A) Methyl region of the MeTROSY spectrum of the nucleosome containing segmentally isotope-labeled DNA-601 with thymine assignments indicated. (C) Overlay of six titration points, containing 0, 1, 2, 3, 4 and 6 equivalents of aclarubicin. (D) The labeled thymines are indicated as cyan spheres in the partial structure of the nucleosome (grey). Arrows indicate outward facing minor grooves with corresponding superhelix location, which are also the aclarubicin binding sites. (E) Bar plot of intensity ratios of the 2, 4 and 6 equivalents aclarubicin titration points compared to the apo spectrum for each thymine methyl peak in the nucleosome (see panel C).

ACLARUBICIN TITRATION

To test the feasibility of using thymine methyl groups as local probes for interaction studies of the nucleosome, the nucleosome-containing segmentally labeled DNA-601 was subjected to a titration with aclarubicin, an anti-cancer drug that intercalates into the DNA minor groove. As it inserts itself between the bases of the DNA, binding will change the chemical environment and thus the chemical shift of the thymine methyl groups (16). The titration of up to 6 equivalents of aclarubicin led to a significant intensity drop of all methyl peaks except for T27 (Fig. 7B). This could be explained by aclarubicin binding in the minor groove of SHL+6, as the proximity of the protons from aclarubicin and rigidification of the DNA due to intercalation would lead to rapid relaxation of the methyl signals, causing them to broaden out. Interestingly, at 6 equivalents, the labeled thymine that is located the furthest inside the nucleosome structure maintains its intensity, which suggests that the next minor groove (SHL+5) is not occupied by an aclarubicin molecule yet (Fig. 7C, D and E). Every equivalent of aclarubicin is likely to bind one minor groove of the nucleosomal DNA, starting from the linker DNA moving inward. As there is also a minor amount of free DNA present in the sample for which aclarubicin has a higher affinity, it is possible that at 6 equivalents SHL+5 has not been reached yet. The titration has not been continued further, as intercalation of aclarubicin at SHL±5 and further would interfere with the bended structure of the nucleosomal DNA. Addition of more aclarubicin would therefore most likely induce nucleosome unfolding and precipitation.

DISCUSSION

In this chapter we explored the feasibility of isotope-labeling DNA for NMR studies of the nucleosome. We described the successful incorporation of uniformly ^{13}C -labeled thymine nucleotides around SHL+6 of the 167 bp DNA-601 sequence and showed that this can be used to probe site-specific effects in the context of the nucleosome. To boost the TROSY effect and obtain higher sensitivity, one could use fully deuterated thymines with specific ^{13}C , ^1H labeling on the methyl group only, similar to the common procedure for methyl-specific labeling in proteins. This would allow for better peak resolution and sensitivity, also after binding of interacting molecules.

A fundamental consideration in the observation of nucleosomal DNA by NMR is the fact that *in vitro* reconstituted nucleosomes always contain a certain amount of free DNA. On the one hand, free DNA (~100 kDa) is likely more sensitively detected than DNA incorporated in nucleosomes (~200 kDa), even if both are very large molecules. On the other hand, the more compact shape of the nucleosome compared to free DNA may boost the effective overall tumbling and thus sensitivity. Furthermore, interacting proteins that bind to DNA in a nucleosome will also bind to free DNA, possibly with even higher affinity, which obscures the interpretation of the observed spectral changes. Free DNA is difficult to remove from nucleosome preparation by ion exchange or size exclusion chromatography, but purification might be achieved by using an affinity tag on one of the histones. Here, native PAGE analysis of our nucleosome preparations indicated presence of <25% free DNA in the NMR sample. Yet, the onset of the intensity decrease and the site-specific differences herein strongly suggest that the observed changes relate to the nucleosomal rather than the free DNA.

In this feasibility study, we chose to label SHL+6, as it is close to the binding site of several interacting proteins at SHL-2 on the neighboring DNA gyre. SHL+6 is fairly close to the end of the DNA sequence, which allows production of the nucleosomal DNA by PCR using a labeled primer. Ideally, the method could be extended to probe other regions of the nucleosomal DNA that are more towards the dyad. Although the use of a 147 bp DNA sequence would already extend the area that can be labeled via primer synthesis, the primer cannot be extended too much, as this would decrease efficiency of the PCR reaction. Alternatively, the nucleosomal DNA could be produced

in two parts and ligated after PCR production of both segments. This enables the incorporation of an isotope-labeled segment anywhere in the sequence, but also adds additional ligation and purification steps to the protocol, making the method less suitable for the production of the large sample quantities required for NMR. However, using a labeled primer up to ~35 nt in a 147 bp sequence can already cover the region from SHL±7 to SHL±4 (depending on the strand that is chosen for labeling). As protein binding to DNA is likely to also affect the chemical shift of the labeled residues in the neighboring DNA gyre, this would hypothetically allow for interaction studies in all areas of the nucleosomal DNA, apart from the central 20 bp, as these do not have a neighboring gyre.

We showed in this study that assignment of the observed thymine resonances is essential to obtain high-resolution information on interactions. However, if assignment is not feasible due to signal overlap, the signals could together also be used as one local probe to report on changes in the region as a whole. Alternatively, incorporation of a paramagnetic label on one of the histones could be considered, using pseudo-contact shifts (PCS) to improve signal dispersion and the structure-dependence of PCS to aid assignment. Changing the chemical shift of the labeled nucleotides by paramagnetic labeling of a histone might also help to discriminate between signals from DNA bound to the nucleosome and free DNA, as the latter would be less affected by the paramagnetic effect.

The results on the aclarubicin-nucleosome interaction strongly suggest that aclarubicin initially intercalates in free linker DNA and progressively inwards into the nucleosomal DNA, to the point where intercalation causes unfolding of the nucleosome and thus effectively eviction of histones. Histone eviction was previously suggested to be caused by a specific competition of aclarubicin with minor groove anchoring arginine in H4 (12). Our data suggest that histone eviction is a more gradual process driven by the intrinsic preference of aclarubicin for intercalation into the free DNA and actuated by the structural rigidity of intercalated DNA. As such, structure-based optimization of anthracyclin compounds such as aclarubicin should rather focus on optimizing this indirect structural effect than aiming to optimize a direct competition with histone-DNA interactions.

In conclusion, we demonstrated here the feasibility of site-specific studies of nucleosomal DNA by NMR. As such, the approach outlined here opens up a route towards more detailed studies of nucleosomal DNA.

MATERIALS AND METHODS

SYNTHESIS OF THE LABELED PRIMER

The isotope-labeled primer was synthesized in a fill-in reaction in Klenow buffer (50 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 1 mM DTT) with 10 μM template and 12.5 μM initial primer (sequences in Table 1). This mixture was heat-annealed for 1 min at 95°C, followed by addition of 50 μM dCTP and dGTP, 100 μM of U-¹³C/¹⁵N-labeled dTTP and dATP as well as 160 units of Klenow fragment (produced in-house) per mL of reaction. The reaction was incubated for 3-4h at RT. The resulting partial DNA duplex was purified and concentrated on a ResourceQ column (GE Healthcare, 1 mL). DNA was eluted by applying a linear gradient from 0-1.5 M NaCl solution over 22 column volumes in 25 mM Tris HCl, pH 8.0, 1 mM EDTA. Eluted DNA was EtOH precipitated at -20°C overnight. DNA pellets were air-dried and dissolved in Milli-Q water.

The purified partial DNA duplex was mixed with gel loading buffer (50 mM Na₂EDTA in formamide, 0.02% bromophenol blue, and 0.02% xylene cyanol) and boiled at 95°C for 2 min before loading on a preparative polyacrylamide gel (10%, 1 x TBE, 8 M urea, 195 x 160 x 1.5 mm). Electrophoresis at 600 V was carried out until the bromophenol blue band reached the bottom of the gel. DNA was visualized using ultraviolet (UV) shadowing and the labeled primer was excised from the gel. Gel slices were crushed by forcing them through a 20 mL syringe. TE-buffer was added (3 mL per 0.5 mL of crushed gel) and the sample was frozen for 30 min at -80°C. After quick thawing at 50°C, the crushed gel was soaked at 90°C for 5 min. DNA was eluted on a rotary shaker overnight at room temperature. Eluted DNA was collected and passed through a 0.22 μm filter, before concentrating it using the ResourceQ column as described above. The EtOH-precipitated pellet was dissolved in Milli-Q water and washed extensively on a centrifugal filter unit to remove residual EtOH. A yield of 6-7 nmol of pure labeled primer per mL of reaction was achieved.

Table 1. Oligonucleotide sequences

Name	Sequence	Purity grade
Template	5'-GTCAGATATATACATCCTGTGCTATCGCAGACAGAGCT-3'	HPLC grade
Initial primer	5'-CTGCGATAGCACAGG-3'	HPLC grade
PCR reverse	5'-CTGGAGCATAATGGAGAATCCCGG-3'	desalt
6T-strand	5'-GATGTATATATCTGA-3'	HPLC-grade
Comp-strand	5'-TCAGATATATACATC-3'	HPLC-grade

LARGE-SCALE PCR PRODUCTION OF SEGMENTALLY ISOTOPE-LABELED DNA601

Segmentally isotope-labeled DNA601 was produced by PCR using 0.8 nM of double-stranded DNA-601 (167 bp) as a template, 1 μ M of labeled primer and 1 μ M of PCR reverse primer (Table 1), 0.2 mM of dNTPs and 8-12 U of Pfu polymerase (produced in-house) per 200 μ L reaction in Pfu buffer (20 mM TrisHCl pH 8.8, 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 2 mM MgSO_4 , 0.1% Triton X-100). Thermocycling was performed in a Biometra TProfessional Standard machine with a 60-well Thermo block to ensure maximum speed and temperature uniformity of the 200 μ L reactions. The PCR program was 2 min of initial melting at 95°C, followed by 23 cycles of 42 s at 95°C, 42 s at 60°C for annealing and 42 s at 72°C for elongation and 5 min of final polymerization at 72°C. Pooled PCR reactions were purified and concentrated on the ResourceQ column, using a salt gradient from 0-1.5 M NaCl over 75 column volumes, holding the salt level as soon as DNA-601 starts eluting. The EtOH-precipitated pooled fractions were dissolved in MQ water and washed extensively on a centrifugal filter unit to remove residual EtOH. Buffer was adjusted to phosphate (20 mM NaPi, pH 7.0, in 100% D₂O) for NMR. A yield of ~0.4 mg from 12 mL PCR (one run 60x 0.2 mL) was achieved.

RECONSTITUTION OF NUCLEOSOMES WITH LABELED DNA-601

Nucleosomes were reconstituted by salt gradient deposition as described before (17). In short, refolded *X. laevis* histone octamers (1.69 mg) were mixed with the labeled DNA-601 (1.61 mg) in 2M KCl buffer (10 mM TrisHCl, pH 7.5, 1 mM EDTA, 1 mM DTT) and gradually dialyzed to 250 mM KCl buffer for 11h at 4°C. The resulting nucleosomes

were spun down to remove precipitate, concentrated and buffer exchanged to phosphate buffer (20 mM NaPi in 100% D₂O, pH7.0) for NMR.

NMR EXPERIMENTS

NMR experiments were carried out on a Bruker Avance III-HD spectrometer at 950 MHz ¹H Larmor frequency, a Bruker Avance III spectrometer at 900 MHz ¹H Larmor frequency (aclarubicin titration), a Bruker Avance III-HD spectrometer at 850 MHz (duplex assignment, labeled partial duplex spectra) or a Bruker Avance III spectrometer at 600 MHz. All magnets were equipped with a TCI cryo-probe. Spectra on pure DNA samples were recorded at 283, 298 or 318 K; all spectra on the NCP were recorded at 318 K.

The ¹³C-filtered NOESY was recorded for the labeled partial duplex (30 + 38 nt) at a concentration of 150 μM in D₂O. The sample was purified on a ResourceQ column and EtOH precipitated. The air-dried pellet was dissolved in MQ and then exchanged to D₂O. Spectra were recorded with a NOESY mixing time of 100 ms.

For assignment of the unlabeled DNA duplex, commercial oligonucleotides (Comp and 6T) were desalted separately using a PD MidiTrap G10 column and then annealed by heating at 95°C for 5 min and slowly cooling down to allow for annealing. The final concentration was ~80 μM of dsDNA in 50 mM NaPi, pH 7.0 in 100% D₂O. Assignments were made based on NOESY and TOCSY spectra recorded at 283 K with mixing times of 200 ms and 80 ms, respectively. Assignment was 100% complete for both strands.

NMR titration experiments on the segmentally labeled NCP were performed using ~24 μM of sample and a 24 μM aclarubicin stock, both in NaPi buffer (20 mM NaPi, pH 7.0) in 100% D₂O. Chemical shift perturbations and intensity changes were monitored by recording constant-time ¹³C-¹H 2D MeTROSSY spectra.

All spectra were processed in NMRPipe or Bruker Topspin and analyzed in Sparky.

CHEMICAL SHIFT PREDICTIONS

Chemical shift predictions for the labeled partial duplex were performed using the DSHIFT web server (<http://www.chem.cuhk.edu.hk/DSHIFT>) (13-15).

REFERENCES

1. Zaret, K.S. and Mango, S.E. (2016) Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev*, **37**, 76-81.
2. Clapier, C.R., Iwasa, J., Cairns, B.R. and Peterson, C.L. (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*, **18**, 407-422.
3. van Nuland, R., van Schaik, F.M., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.M. and van Ingen, H. (2013) Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin*, **6**, 12.
4. Musselman, C.A., Gibson, M.D., Hartwick, E.W., North, J.A., Gatchalian, J., Poirier, M.G. and Kutateladze, T.G. (2013) Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun*, **4**, 2969.
5. Gatchalian, J., Wang, X., Ikebe, J., Cox, K.L., Tencer, A.H., Zhang, Y., Burge, N.L., Di, L., Gibson, M.D., Musselman, C.A. *et al.* (2017) Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nat Commun*, **8**, 1489.
6. Morrison, E.A., Bowerman, S., Sylvers, K.L., Wereszczynski, J. and Musselman, C.A. (2018) The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife*, **7**.
7. Zhou, C.Y., Johnson, S.L., Gamarra, N.I. and Narlikar, G.J. (2016) Mechanisms of ATP-Dependent Chromatin Remodeling Motors. *Annu Rev Biophys*, **45**, 153-181.
8. Louis, J.M., Martin, R.G., Clore, G.M. and Gronenborn, A.M. (1998) Preparation of uniformly isotope-labeled DNA oligonucleotides for NMR spectroscopy. *J Biol Chem*, **273**, 2374-2378.
9. Kato, H., van Ingen, H., Zhou, B.R., Feng, H., Bustin, M., Kay, L.E. and Bai, Y. (2011) Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A*, **108**, 12283-12288.
10. Mueller-Planitz, F., Klinker, H. and Becker, P.B. (2013) Nucleosome sliding mechanisms: new twists in a looped history. *Nat Struct Mol Biol*, **20**, 1026-1032.
11. Farnung, L., Vos, S.M., Wigge, C. and Cramer, P. (2017) Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature*, **550**, 539-542.
12. Pang, B., Qiao, X., Janssen, L., Velds, A., Groothuis, T., Kerkhoven, R., Nieuwland, M., Ovaa, H., Rottenberg, S., van Tellingen, O. *et al.* (2013) Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat Commun*, **4**, 1908.
13. Altona, C., Faber, D.H. and Hoekzema, A.J.A.W. (2000) Double-helical DNA ¹H chemical shifts: an accurate and balanced predictive empirical scheme. *Magnetic Resonance in Chemistry*, **38**, 95-107.
14. Lam, S.L. (2007) DSHIFT: a web server for predicting DNA chemical shifts. *Nucleic Acids Res*, **35**, W713-717.

15. Wijmenga, S.S., Kruithof, M. and Hilbers, C.W. (1997) Analysis of (1)H chemical shifts in DNA: Assessment of the reliability of (1)H chemical shift calculations for use in structure refinement. *J Biomol NMR*, **10**, 337-350.
16. Yang, D. and Wang, A.H. (1994) Structure by NMR of antitumor drugs aclacinomycin A and B complexed to d(CGTACG). *Biochemistry*, **33**, 6595-6604.
17. Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M. and Luger, K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol*, **375**, 23-44.

SUPPLEMENTARY INFORMATION

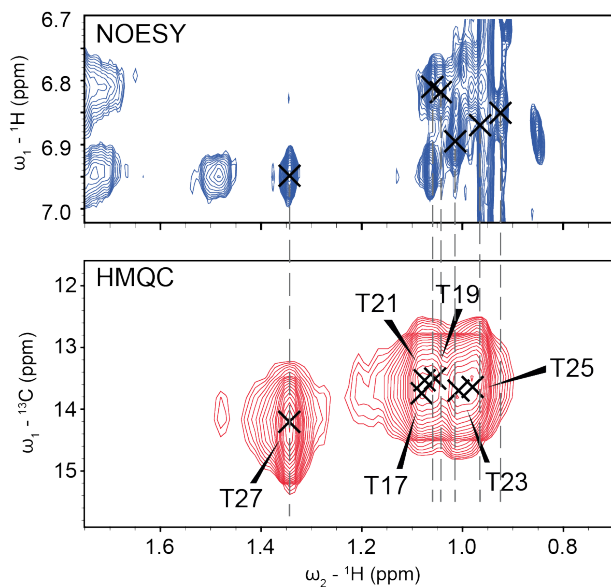


Figure S1. Methyl group assignment transfer from NOESY sequential walk (unlabeled duplex, top panel) to the HMQC of the labeled partial duplex (bottom panel).

GENERAL DISCUSSION

Thinking about genomic DNA, the first process that comes to mind might be transcription, representing the first step in the central dogma of molecular biology. This process is usually visualized as taking place on naked DNA. However, in reality, genomic DNA is packaged in chromatin. Since the DNA is partially occupied by nucleosomes, this packaging has profound implications for the mechanism of transcription and other DNA-templated processes. Structural and mechanistic studies are now starting to shed light on the molecular details involved, as in the passage of RNA polymerase through chromatin (1-3).

Apart from processes like transcription and replication that operate on the DNA directly, many other aspects of chromatin biology depend in some way on the DNA. At the simplest level, this becomes clear when looking at structures of the nucleosome where histone tails interact with the DNA and mediate compaction of nucleosomes (4,5). Also, nucleosomes can breathe, transiently exposing more of their DNA (6,7). In epigenetic regulation of nucleosome function, many of the nucleosome-binding proteins that target a specific binding site on the histones to place, bind or erase post-translational modifications, often bind additionally to nucleosomal DNA. While these interactions are often non-specific ionic interactions, they can critically modulate binding affinity and specificity and alter nucleosome stability (8-10). Finally, chromatin remodelers depend on the deformability of the DNA and DNA-histone interactions to alter the chromatin landscape.

Given the crucial importance of nucleosomal DNA in nucleosome function, there is a need to focus structural studies not only on the histones but also on the DNA. This is challenged in part by the need for high resolution structures to resolve the molecular details of DNA structure in the nucleosomes (11) as well as the key role for DNA deformability and dynamics that is now emerging. In this thesis, we focused on several aspects of the nucleosomal DNA, from fundamental design principles of the nucleosome, and the role of DNA in activation of a remodeler to development of an

efficient approach to produce genomic nucleosomal DNA and new techniques to unlock the nucleosomal DNA for NMR studies on the nucleosome.

DNA VERSUS HISTONES IN NUCLEOSOME FUNCTION

Despite the importance of DNA in nucleosome function, one could argue that DNA is a much less interesting factor than the histone proteins. DNA has only four building blocks on offer, as opposed to 20 amino acids forming proteins. The histones have evolved from being basic DNA-compacting proteins in archaeal hypernucleosomes to being highly variable epigenetic interaction platforms in eukaryotic nucleosomes, providing specificity and regulation of DNA-templated processes. The DNA has stayed largely the same over all this time. As an interaction partner, DNA is much less complex than the histones, its main feature being the extensive negative charge on the exposed backbone, and it is less diversely modified than histones. In addition, it has been shown that the structure of the nucleosome barely varies with different DNA sequences, leaving no room for differentiation of interactions in the tertiary conformation either. So, exactly how does DNA contribute to nucleosome function?

The distinct feature of DNA is of course the sequence of the bases. While there are only four options for the bases, the conformation of the DNA backbone is determined by six main torsion angles, while an additional sixteen rigid-body parameters, including base-pair and base-pair-step parameters, determine the conformation of the bases inside the helix (12). This makes DNA actually more complex than proteins, as these have an overall fold that is determined by only two backbone angles. It also means that DNA conformation can be subtly altered in many ways. Such sequence-specific deformation of DNA is thought to be essential for some transcription factor binding (13). It may also be critical for the nucleosome's altering response to remodelers or transcription factor binding.

Because of the highly bent conformation of DNA in the nucleosome, the stability and likely also the dynamics of the nucleosome are sequence-dependent. This dependency was one of the challenges to overcome during the first attempts to crystallize the nucleosome and remains a challenge for *in vitro* nucleosome reconstitution. Strong-positioning sequences yield the most stable and uniform nucleosomes. *In vivo*, however, nucleosomes will be assembled on a wide range of

DNA sequences resulting potentially in more open and more dynamic structures with higher availability of the DNA for interactions or higher potential for remodeling. Data obtained with strong-positioning sequences may thus reflect only part of the conformational landscape accessible to genomic nucleosomes. Finally, the nucleosomal DNA sequence can harbor sequence motifs for transcription factor binding. Although most transcription factors target only free DNA, pioneer factors are able to recognize their target motifs even if these are incorporated into nucleosomes. To understand the true nature of nucleosomes and the mechanism of remodeling and pioneer factor interactions, it is thus crucial to have a detailed view on the DNA within the nucleosome.

DESIGN PRINCIPLES OF THE NUCLEOSOME

In **Chapter 2** we built a de novo model of the archaeal hypernucleosome, on the basis of the eukaryotic nucleosome and archaeal histone structures. Since the hypernucleosome is formed from a single histone species, the structure is symmetric and exploiting this symmetry was key to building this model. Starting from an ideal DNA superhelix, histone dimers were assembled onto the DNA guided by the histone dimer contacts with the DNA and in particular, the arginine residues anchoring into the inward facing DNA minor grooves at every turn. Notably, use of the eukaryotic nucleosomal DNA conformation, which is heavily stretched and kinked, failed to result in a proper hypernucleosome conformation, illustrating that the DNA in eukaryotic nucleosomes is adapted to the irregular shape of the histone heterodimers in the eukaryotic nucleosome.

Comparing homology models of other archaeal histone-DNA complexes, we concluded that the surface shape and symmetry of the dimer determine the type of nucleosome-like complex that can be formed. Dimer symmetry ensures identical interfaces between subsequent dimers, facilitating multimerization. This also demonstrated why the asymmetric eukaryotic histone dimers are not able to form hypernucleosomes. Furthermore, we hypothesize that stacking interactions in the hypernucleosome influence the stability of the complex and thereby also its compactness. This might represent a mechanism of regulation for protein interactions and gene expression in archaeal chromatin, a role that is fulfilled by histone tails and post-translational modifications in eukaryotic chromatin.

Studying DNA structure and interactions by NMR has always been challenging for several reasons. The low number of building blocks is a major disadvantage as it causes extensive signal overlap in the spectra. This problem increases with DNA length, and the corresponding higher molecular weight also causes line broadening, lowering the resolution that can be reached. Although NMR has been used to study structural elements in a 100 kDa RNA molecule (14), the largest solution NMR structure of a double-stranded DNA molecule that has been determined to date is 21 bp long (15), illustrating the challenge of studying full-length nucleosomal DNA of ~147 bp. Pilot experiments in our lab showed that the NMR spectra of nucleosomal DNA are dominated by the flexible termini, further limiting the insight that can be obtained. As a consequence, usually only the histones are isotope-labeled and observed in NMR studies on nucleosomes, and the role of the 'invisible' DNA might be easily overlooked. Is NMR then really the most suitable method to study it? Other high-resolution techniques such as X-ray crystallography and cryoEM have been able to capture the DNA inside the nucleosome in atomic detail, also in complex with interacting proteins. However, these methods only give a static view on the complexes in a non-native environment. Using Förster resonance energy transfer (FRET), the dynamics of DNA in a nucleosomal context can be captured, but NMR is the only high-resolution method that can give both structural and dynamic information at atomic resolution in solution. In this thesis, we therefore explored the possibilities of high-resolution NMR studies on nucleosomal DNA and interactions involving nucleosomes.

NEW METHODOLOGIES FOR DNA PRODUCTION AND ISOTOPE-LABELING

To enable NMR studies on nucleosomal DNA and genomic nucleosomes, new methods were developed. In particular, such studies require strategies for efficient large-scale production of DNA of a specific sequence and for specific isotope labeling of DNA.

Several methods are available for large-scale production of DNA, including plasmid-based production in *E. coli*, large-scale PCR and ramified RCA (described in **Chapter 3** of this thesis). Which method to choose depends on the desired sequence and labeling scheme. Synthesis of the strong-positioning DNA-601 sequence that is

regularly used in structural studies is most commonly done using the plasmid-based method. In **Chapter 3**, we showed that rRCA is an efficient alternative with the major advantage that it is less labor-intensive, has higher yield and is more flexible in choice of sequence. The sequence flexibility offered by rRCA opens up opportunities to use native genomic DNA sequences for nucleosome reconstitution. We demonstrated that genomic nucleosomes as occurring on the 3' end of the *LIN28B* locus in chromosome 6 can be reconstituted. This opens up an avenue towards studies of both the dynamic properties of this nucleosome, as well as binding of the OSKM pioneer transcription factors that drive reprogramming of somatic cells into iPSCs.

The efficiency of the method could be enhanced by performing rRCA and digestion simultaneously, using a methylated circular template and a methylation-sensitive restriction enzyme. This might reduce the total reaction time and increase product formation due to lowered viscosity. Longer sequences, like DNA for nucleosome arrays can best be synthesized using the plasmid-based method or rRCA, as these have a lower error rate for long sequences than PCR. However, for both methods the cloning of a multi-repeat template for alternative sequences is a challenge.

Uniform isotope-labeling of nucleosomal DNA is ideally done by plasmid production, as isotope-labeled nucleotides used in PCR and rRCA are costlier than ^{13}C -glucose and $^{15}\text{NH}_4\text{Cl}$ used in the cell culture medium. However, as uniform labeling does not resolve the issues of peak overlap and line broadening in NMR spectra of the ~100 kDa nucleosomal DNA, site-specific labeling is required to achieve a detailed view on the DNA. For segmental one-strand isotope-labeling of a specific set of thymines, as described in **Chapter 5** of this thesis, the only suitable method is PCR, as both plasmid production and rRCA are essentially one-pot reactions without the requirement of each repeat being initiated by an individual primer that contains the isotope-labeled nucleotides. To reduce the workload of doing 4000 PCR reactions for one NMR sample, for this synthesis a Thermocycler was used that can handle larger tubes, increasing the volume per plate from 4.8 mL to 12 mL. Although this setup makes PCR also more feasible for large-scale synthesis of unlabeled DNA of alternative sequences, rRCA is still more convenient in such cases, mostly because it is a one-pot isothermal reaction that can be easily scaled up. This makes aliquoting of reaction volumes for thermocycling redundant and therefore strongly reduces the workload.

In **Chapter 5**, we describe that isotope-labeled thymine methyl groups can provide site-specific information on the binding of the DNA-intercalator aclarubicin. This method can be extended to probe different regions of the nucleosomal DNA, making it suitable for the study of interactions with other molecules or proteins as well. However, for many DNA-interacting proteins, the position of these probes on the base inside the helix, might not be ideal to detect changes. Although the naturally abundant ^{31}P in the DNA backbone might be most directly in contact with interaction partners, using this nucleus allows to observe the flexible DNA ends only and resolution in these spectra is very poor. Proton is the most sensitive nucleus for NMR and has a much better peak dispersion. Therefore, alternative probes might be the protons on the sugar ring, that are closer to the backbone and the interaction surface of the DNA than the protons on the base. While the sugar protons may be less sensitive than the thymine methyl probes in the methylTROSY approach, combination with deuteration and selectively isotope-labeled nucleotides will allow the use of TROSY to boost sensitivity. The use of solid-state NMR might enable the observations of the sugar carbons as local probes, although also there, signal overlap will be a challenge. An alternative strategy could employ paramagnetic labeling of DNA to indirectly observe nucleosome interactions from the DNA perspective. Site-specific incorporation of such a label is only possible via incorporation of a primer, limiting the placement within the DNA. In addition, as paramagnetic labels are usually large, care must be taken that placement of these tags does not interfere with the interaction of interest. At the same time, a paramagnetic ion of sufficient strength should be chosen as to still observe the effect of it on the binding partner.

PERSPECTIVE ON ISWI REMODELING

Nucleosome remodeling is one of the most complex processes taking place on the nucleosome, having the DNA as a primary interaction target. Although more and more structural information has become available on ISWI and other remodelers, the molecular mechanism is still subject to different interpretations. Defining remodeling as the process of ATP-fueled movement of a nucleosome along the DNA, several key aspects can be identified to decipher with regards to the mechanism. Firstly, the question rises what the energy from ATP hydrolysis is actually used for. Intuitively, it would make sense that ATP hydrolysis directly provides the power needed to pull on

the DNA and move it to its next position. Alternatively, the energy released by ATP hydrolysis might be used to reset the structure of the remodeler to its original conformation, preparing it for the next movement. This leads to another key question concerning the mechanism of remodeling: is the motion of the lobes that causes movement of the DNA intrinsically encoded in the conformational space of the protein? If it is, this would argue in favor of the idea that ATP is used to convert the remodeler from a post-translocation state back to the pre-translocation state, comparable to stretching a spring before releasing it again. If it is not, that means that ATP hydrolysis would directly provide the energy needed to rotate the lobes with respect to each other, thereby orchestrating the translocation of the DNA.

In **Chapter 4**, we describe our NMR-driven approach to characterize the release of autoinhibition of the ISWI remodeler. ISWI is one of the smallest ATP-dependent remodelers but nevertheless contains the conserved ATPase domains, making it a suitable model for NMR studies. Many studies before having described the H4 tail to be the primary actor in release of autoinhibition by the NTR, our studies show that the H4 tail can only bind to ISWI after addition of DNA. We hypothesize that the DNA has a crucial role in loosening the basic NTR to increase the chances of the H4 tail to bind and stabilize the active state in context of a nucleosome. This illustrates that, although the DNA is very abundant in the cell nucleus and does not feature any specific binding epitope, the combination of DNA and histone binding can lead to a specific and regulated binding event. The transition to the active state of ISWI after release of autoinhibition occurs by a rotational movement of Lobe 1 with respect to Lobe 2 to bring the helicase motifs in close proximity. This large conformational change suggests that ISWI is strongly strained in the autoinhibited state and returns to a more energetically favorable conformation when the NTR no longer inhibits this transition and becomes locked once the DNA binding site finds the nucleosomal DNA. We speculate that the conformational change triggered by DNA and H4 tail binding and the release of the strain (partially) initiates the first translocation event, possibly aided by the intrinsic conformational landscape of the ATPase domain. This would mean that ATP is thus required to reset the spring and convert the lobes back to their pre-translocation state.

In the last year, several cryoEM structures became available of ISWI bound to the nucleosome in different activation states, which contributed significantly to our

understanding of the mechanism of remodeling. However, the dynamic regions, such as the autoregulatory domains, are usually not well resolved using this method. NMR is much more suitable to study the dynamic aspects of remodeling, but full-length ISWI and even the isolated ATPase domain is quite large for solution NMR and difficult to maintain at the high concentrations that are required. Future NMR studies might therefore use deuterated and ILV-labeled ISWI to enable the use of MeTROSY experiments, which could enhance sensitivity and resolution. This would allow the observation of signals from throughout the protein, including dynamic and core regions and therefore potentially yield more information on conformational changes during activation. Also truncated constructs and mutagenesis could increase the number of assigned residues and thereby expand the amount of site-specific information that can be obtained.

Finally, it would be exciting to combine site-specific labeling of nucleosomal DNA with ISWI binding to directly observe this interaction from the DNA perspective by NMR. This would require full deuteration of ISWI and the histones close to the binding site to reduce relaxation in the 275 or even 350 kDa complex of the nucleosome with one or two remodelers bound. Although chemical shift perturbation might be difficult to observe due to the thymine methyl groups being buried inside the double helix, line broadening upon binding of ISWI might reveal dynamical changes in the DNA that are difficult to pick up by X-ray or cryoEM studies. As suggested before, the use of the sugar protons as probes, or the sugar carbons in a ssNMR approach might expand the amount of information that can be obtained. This could yield new information on dynamics of the DNA after ISWI binding and in presence or absence of nucleosomal H4 tail and/or ADP or ATP, adding another dimension to our understanding of the molecular mechanism of ISWI remodeling.

CONCLUSION

The studies described in this thesis aimed to contribute to our basic understanding of chromatin structure and function, involving chromatin remodeling and pioneer factor binding in particular. To obtain a fundamental understanding of nucleosome architecture and higher order chromatin, we used a modeling approach to build an archaeal hypernucleosome, the evolutionary predecessor of the eukaryotic nucleosome. Furthermore, we developed multiple strategies to synthesize nucleosomal DNA with different sequences and isotope-labeling requirements for NMR studies on nucleosome-protein interactions, including pioneer factors and nucleosome remodelers. These methods may serve to explore new approaches in studying the structural basis of nucleosome interactions and thereby expand our knowledge of DNA-templated processes such as gene expression and DNA repair, and understand how these are regulated in health and disease.

REFERENCES

1. Ehara, H., Kujirai, T., Fujino, Y., Shirouzu, M., Kurumizaka, H. and Sekine, S.I. (2019) Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science*, **363**, 744-747.
2. Farnung, L., Vos, S.M. and Cramer, P. (2018) Structure of transcribing RNA polymerase II-nucleosome complex. *Nat Commun*, **9**, 5432.
3. Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.I. and Kurumizaka, H. (2018) Structural basis of the nucleosome transition during RNA polymerase II passage. *Science*, **362**, 595-598.
4. Kan, P.Y., Lu, X., Hansen, J.C. and Hayes, J.J. (2007) The H3 tail domain participates in multiple interactions during folding and self-association of nucleosome arrays. *Mol Cell Biol*, **27**, 2084-2091.
5. McBryant, S.J., Klonoski, J., Sorensen, T.C., Norskog, S.S., Williams, S., Resch, M.G., Toombs, J.A., 3rd, Hobdey, S.E. and Hansen, J.C. (2009) Determinants of histone H4 N-terminal domain function during nucleosomal array oligomerization: roles of amino acid sequence, domain length, and charge density. *J Biol Chem*, **284**, 16716-16722.
6. Buning, R., Kropff, W., Martens, K. and van Noort, J. (2015) spFRET reveals changes in nucleosome breathing by neighboring nucleosomes. *J Phys Condens Matter*, **27**, 064103.
7. Li, G. and Widom, J. (2004) Nucleosomes facilitate their own invasion. *Nat Struct Mol Biol*, **11**, 763-769.
8. Musselman, C.A., Gibson, M.D., Hartwick, E.W., North, J.A., Gatchalian, J., Poirier, M.G. and Kutateladze, T.G. (2013) Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun*, **4**, 2969.
9. Richart, A.N., Brunner, C.I., Stott, K., Murzina, N.V. and Thomas, J.O. (2012) Characterization of chromoshadow domain-mediated binding of heterochromatin protein 1alpha (HP1alpha) to histone H3. *J Biol Chem*, **287**, 18730-18737.
10. van Nuland, R., van Schaik, F.M., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.M. and van Ingen, H. (2013) Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin*, **6**, 12.
11. Richmond, T.J. and Davey, C.A. (2003) The structure of DNA in the nucleosome core. *Nature*, **423**, 145-150.
12. Lu, X.J. and Olson, W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc*, **3**, 1213-1227.
13. Perino, M., van Mierlo, G., Karemaker, I.D., van Genesen, S., Vermeulen, M., Marks, H., van Heeringen, S.J. and Veenstra, G.J.C. (2018) MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat Genet*, **50**, 1002-1010.

14. Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E.L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A. *et al.* (2011) NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*, **334**, 242-245.
15. Masliah, G., Rene, B., Zargarian, L., Femandjian, S. and Mauffret, O. (2008) Identification of intrinsic dynamics in a DNA sequence preferentially cleaved by topoisomerase II enzyme. *J Mol Biol*, **381**, 692-706.

ENGLISH SUMMARY

Chromatin provides the required structural compaction of DNA to fit in the nucleus and plays crucial roles in controlling cell fate and protecting genome integrity. Chromatin function ultimately depends on the structural and dynamical properties of the nucleosome, its fundamental repeating unit. The nucleosome acts as a gate keeper by regulating the binding of proteins that carry out DNA-templated processes like transcription, replication and repair. The studies described in this thesis aimed to contribute to our basic understanding of chromatin structure and function, involving chromatin remodeling and pioneer factor binding in particular. We specifically focused on the contribution of the nucleosomal DNA, ranging from fundamental design principles of the nucleosome, and the role of DNA in activation of a remodeler to development of an efficient approach to produce genomic nucleosomal DNA and new techniques to unlock the nucleosomal DNA for NMR studies on the nucleosome.

ES

In **Chapter 1**, we reviewed the contribution of NMR to our understanding of the molecular basis of nucleosome function. Starting from the pioneering NMR studies in the 1960's, we focused on the progress in sample preparation and NMR methodology that has allowed high-resolution studies on the nucleosome and its subunits. We summarized the results and approaches of state-of-the-art NMR studies on nucleosomal DNA, histone complexes, nucleosomes and nucleosomal arrays. This highlighted the particular strength of NMR in studying nucleosome dynamics and nucleosome-protein interactions. Looking ahead, we discussed the exciting new possibilities that will be afforded by on-going developments in solution and solid-state NMR. By increasing both the depth and breadth of nucleosome NMR studies, it will be possible to offer a unique perspective on the dynamic landscape of nucleosomes and its interacting proteins.

In **Chapter 2**, we used computational modeling to study newly discovered archaeal histones that were only identified at the sequence level. We analyzed their propensity to form a continuous histone-DNA complex, or hypernucleosome, analogous to what was recently reported for archaeal histones HMfA and HMfB. As eukaryotes are classified as being part of the archaeal evolutionary branch, this complex might represent the evolutionary ancestor of chromatin organization into nucleosomes in

eukaryotes. Sequence analysis and structural modeling suggested that the shape, symmetry and surface residues of the histones dictate the type of nucleosome-like complex that can be formed with genomic DNA, as well as its stability and level of compaction. In addition, modeling of archaeal histone complexes highlighted the fundamental design principles of the eukaryotic nucleosome, demonstrating why eukaryotic histones are not able to form a hypernucleosome. We proposed that different archaeal histone dimers in the same organism can specifically promote or disturb hypernucleosome formation and that this might be functional in gene regulation. Differentiation of histones into asymmetric dimers and the emergence of modifiable N-terminal tails may have led to the evolution of eukaryotic octameric nucleosomes and an epigenetic code providing extensive regulation of DNA-templated processes.

In **Chapter 3**, we introduced the method of ramified rolling circle amplification (rRCA) for the efficient large-scale production of a genomic nucleosomal DNA sequence from the human LIN28B locus. We showed that rRCA can be used to produce milligram amounts of nucleosomal DNA in an isothermal, one-pot reaction overnight that is easily scalable. In comparison to the currently used methods, it yields 10-fold more product than PCR, and rivals production using plasmids. Importantly, rRCA offers flexibility in choice of sequence. This may facilitate future studies on native, genomic nucleosomes, which might have structural and dynamic properties different from nucleosomes reconstituted on the common artificial strong-positioning sequences. In addition, it allows to incorporate pioneer transcription factor recognition motifs and to study their binding to the nucleosome. To illustrate this, we showed that rRCA-produced LIN28B DNA forms functional nucleosomes capable of binding pioneer transcription factor Oct4.

Chromatin remodelers are ATP-dependent molecular motors that are essential for proper control of gene expression and protection of genome integrity. In **Chapter 4**, we applied Fluorescence Anisotropy (FA) and Nuclear Magnetic Resonance (NMR) to identify and characterize conformational changes of the ISWI remodeler upon interaction with its activating nucleosomal epitopes. To prevent spurious consumption of ATP, these lobes are locked in an inactive conformation in the free state of the remodeler, and this auto-inhibition is only relieved by interaction with the substrate, in particular the nucleosomal DNA and histone H4 tail. We showed that the ISWI ATPase

domain is only able to bind the H4 tail in presence of dsDNA and that this ISWI-DNA-H4 complex requires presence of ATP-mimic ADP-BeFx for stability. We found that the N-terminal region (NTR) of ISWI, which locks the ATPase lobes in their inactive conformation, is not stable as an isolated domain and likely to be bound to the ISWI core at all stages of activation. Furthermore, our data indicated that dsDNA binding alone induces a conformational change in ISWI with an additional change upon H4 tail binding. Based on these data, we proposed that the first step of autoinhibition release is the interaction of ISWI with dsDNA. In this model, basic residues in the NTR mediate non-specific binding to DNA thereby loosening binding to the ISWI core and thus allowing for a conformational change of the ATPase lobes. We further hypothesized that, once ISWI is bound to the nucleosome in proximity of the H4 tail, this tail binds to the second lobe to stabilize the active conformation and promote further relief of autoinhibition by the NTR.

As the nucleosomal DNA plays a crucial role in nucleosome interactions with chromatin remodelers, pioneer transcription factors and other proteins, in **Chapter 5**, we explored the possibility of directly observing the nucleosomal DNA in NMR studies. We developed a method for segmental one-strand isotope-labeling of six thymine residues in the nucleosomal DNA-601 sequence. The limited number of labeled residues reduces signal overlap and the labeling of thymines, which contain a methyl group, creates the opportunity of using MeTROSY NMR experiments to improve signal to noise ratios in large complexes like the nucleosome. We described the synthesis of the segmentally isotope-labeled DNA601 in NMR sample quantities by large-scale PCR as well as assignment of the six methyl signals. After incorporation of the labeled DNA into nucleosomes, we performed a titration with aclarubicin, an anti-cancer drug that intercalates into the DNA minor grooves, showing that the labeled thymines can serve as local probes to observe site-specific binding events in the context of the nucleosome. This opens up new opportunities to study the contribution of nucleosomal DNA to nucleosomal interactions in molecular detail.

In **Chapter 6**, a critical discussion is provided on the studies presented in this thesis, as well as a broader perspective on the findings and potential future directions to study the structural basis of nucleosome interactions and understand the role of the nucleosomal DNA in these.

NEDERLANDSE SAMENVATTING

Het DNA in de celkern is compact opgevouwen in de vorm van chromatine, wat er onder meer voor zorgt dat DNA in de celkern past. Chromatine speelt daarnaast een essentiële rol in het reguleren van celdifferentiatie en het in stand houden van de integriteit van het genoom. Deze en andere functies van chromatine zijn afhankelijk van de fundamentele eigenschappen van de basiseenheid van chromatine: het nucleosoom. De structurele en dynamische eigenschappen van nucleosomen reguleren de binding van eiwitten die betrokken zijn bij het uitvoeren van transcriptie, replicatie en het repareren van DNA-schade. De studies beschreven in dit proefschrift hadden als doel bij te dragen aan ons fundamentele begrip van de structuur en de functie van chromatine, en in het bijzonder van het verplaatsen van nucleosomen, het zogenaamde *chromatine remodeling*, en de binding van pionier-transcriptiefactoren. Onze focus lag hierbij op de rol van het nucleosomaal DNA. We onderzochten de basale ontwerpprincipes van het nucleosoom en de rol van DNA in de activatie van een remodeler-eiwit. Daarnaast ontwikkelden we een efficiënte methode om genomisch nucleosomaal DNA te produceren en nieuwe technieken om de studie van DNA in het nucleosoom met NMR mogelijk te maken.

In **Hoofdstuk 1** presenteren we een overzicht van de bijdrage van kernspinresonantie (nuclear magnetic resonance, NMR) aan onze kennis van de moleculaire basis van de functie van het nucleosoom. Beginnend bij de baanbrekende NMR-studies in de jaren '60, benadrukken we dat het bestuderen van het nucleosoom in hoge resolutie mogelijk is geworden door grote ontwikkelingen in het prepareren van samples en NMR-methodologie. We beschrijven de belangrijkste resultaten en gebruikte methoden van *state-of-the-art* NMR-studies aan nucleosomaal DNA, histoncomplexen, afzonderlijke nucleosomen en langere strengen van nucleosomen. Dit laat zien dat NMR in het bijzonder geschikt is voor het bestuderen van dynamische eigenschappen van nucleosomen en nucleosoom-eiwit interacties. We kijken vooruit naar de ontwikkelingen in vloeistof NMR en vaste stof NMR die nieuwe mogelijkheden met zich mee zullen brengen. De NMR-studies die met behulp van deze ontwikkelingen mogelijk worden, zouden een uniek, nieuw perspectief kunnen werpen op de dynamische eigenschappen van nucleosomen en de eiwitten die ermee interacteren.

Hoofdstuk 2 beschrijft een aantal structuurmodellen van archaeale histonen waarvan alleen de sequentie geïdentificeerd is. We hebben geanalyseerd hoe waarschijnlijk het is dat deze eiwitten multimeren kunnen vormen op genomisch DNA, vergelijkbaar met de recent gepubliceerde structuur van de archaeale histonen HMfA en HMfB. Omdat eukaryoten deel uitmaken van de archaeale tak in de fylogenetische stamboom, is dit zogenaamde hypernucleosoomcomplex mogelijk de evolutionaire voorganger van chromatine-organisatie met behulp van nucleosomen, zoals we die kennen in eukaryoten. Analyse van de sequenties en structuurmodellen suggereert dat de vorm, symmetrie en de oppervlakteresiduen van de histonen bepalen welk type complex gevormd kan worden met genomisch DNA. Daarnaast zeggen deze eigenschappen mogelijk iets over de stabiliteit van het complex en de mate van compactheid. Verder werd door het modelleren van deze archaeale histoncomplexen ook duidelijk waarom eukaryote histonen niet in staat zijn om hypernucleosomen te vormen en wat dus de fundamentele ontwerpprincipes van het eukaryote nucleosoom zijn. We veronderstellen dat verschillende archaeale histondimeren in hetzelfde organisme de vorming van hypernucleosomen specifiek kunnen bevorderen, dan wel verstoren, en dat dit functioneel kan zijn in genregulatie. Differentiatie van histonen naar asymmetrische dimeren en het ontstaan van modificeerbare N-terminale staarten hebben mogelijk geleid tot de evolutie van de eukaryote octamerische nucleosomen en een epigenetische code die uitgebreide regulatie van DNA-gekoppelde processen mogelijk maakt.

Hoofdstuk 3 introduceert *ramified rolling circle amplification* (rRCA) als methode voor de efficiënte productie van een genomische nucleosomale DNA-sequentie van de humane LIN28B locus op grote schaal. We hebben laten zien dat het met rRCA mogelijk is om overnacht milligram hoeveelheden nucleosomaal DNA te produceren in een isotherme reactie die makkelijk op te schalen is. rRCA heeft een opbrengst die tien keer hoger is dan die van PCR en kan concurreren met de opbrengst uit plasmideproductie. Een belangrijk voordeel van de rRCA methode is dat deze flexibel is in de keuze van de DNA-sequentie. Dit is belangrijk voor toekomstige studies aan genomische nucleosomen, waarvan de structurele en dynamische eigenschappen kunnen verschillen van nucleosomen die gereconstitueerd zijn met de nu gebruikelijke artificiële DNA-sequenties. Daarnaast biedt dit de mogelijkheid voor het incorporeren van herkenningsmotieven van pionier-transcriptiefactoren om hun binding aan

nucleosomen te onderzoeken. Dit illustreren we door te laten zien dat het LIN28B DNA, geproduceerd met rRCA, functionele nucleosomen kan vormen waaraan de pionier-transcriptiefactor Oct4 kan binden.

Chromatine remodeler-eiwitten zijn moleculaire motortjes die voor hun energie afhankelijk zijn van de hydrolyse van ATP. Ze zijn essentieel voor de regulatie van genexpressie en het in stand houden van de integriteit van het genoom. In **Hoofdstuk 4** gebruiken we fluorescentie anisotropie (FA) en NMR voor het identificeren en karakteriseren van conformationele veranderingen van de ISWI remodeler na interactie met activerende epitopen. Om nutteloze consumptie van ATP te voorkomen, zijn de twee helften van het ATPase domein vastgezet in hun inactieve vorm wanneer de remodeler niet op een nucleosoom zit. Deze autoinhibitie wordt opgeheven als de remodeler bindt aan het nucleosomaal DNA en de histonstaart van histon H4. In dit hoofdstuk laten we zien dat het ISWI ATPase domein pas kan binden aan de H4-staart als er dubbelstrengs DNA aanwezig is in de oplossing en dat het ISWI-DNA-H4 complex alleen stabiel is in aanwezigheid van de ADP-BeFx, een stof die erg lijkt op ATP. We hebben gevonden dat de N-terminale regio (NTR) van ISWI, die de ATPase-helften in hun inactieve conformatie vasthoudt, niet stabiel is als geïsoleerd domein. Dit maakt het aannemelijk dat de NTR in alle stadia van activatie in ieder geval deels aan het ISWI ATPase domein gebonden is. Verder laten onze data zien dat de binding van dubbelstrengs DNA op zichzelf al een conformationele verandering van ISWI veroorzaakt, terwijl er nog een extra verandering optreedt na binding van de H4-staart. Gebaseerd op deze data, stellen we voor dat de eerste stap om de autoinhibitie op te heffen bestaat uit de interactie met het DNA. In deze hypothese zijn de basische residuen in de NTR verantwoordelijk voor specifieke binding aan het DNA. Dit zorgt ervoor dat de binding van de NTR aan het ATPase domein verzwakt wordt, waardoor de conformationele verandering van de ATPase-helften kan plaatsvinden. We veronderstellen verder dat zodra ISWI gebonden is aan het nucleosoom in de nabijheid van de H4-staart, deze staart de actieve conformatie stabiliseert en daarmee de verdere opheffing van autoinhibitie door de NTR stimuleert.

Omdat nucleosomaal DNA een belangrijke rol speelt in nucleosoominteracties met chromatine remodelers, pionier-transcriptiefactoren en andere eiwitten, onderzochten we in **Hoofdstuk 5** de mogelijkheid om het nucleosomaal DNA direct te observeren in

NMR-experimenten. Hiervoor hebben we een methode ontwikkeld voor het isotoop-labelen van slechts zes thymine residuen in een specifieke regio en op één streng van de nucleosomale DNA-601 sequentie. Het beperkte aantal gelabelde residuen vermindert de overlap van signalen in het NMR-spectrum. Daarnaast zorgt het labelen van methyl-bevattende thymines ervoor dat MeTROSY experimenten gebruikt kunnen worden om de signaal-ruisverhouding te verbeteren in grote complexen zoals het nucleosoom. We beschrijven de synthese van het gedeeltelijk isotoop-gelabelde DNA601 in hoeveelheden die voldoende zijn voor een NMR-sample met behulp van grootschalige PCR en de toekenning van de zes methylsignalen in het NMR-spectrum. Vervolgens hebben we een titratie uitgevoerd van het in nucleosomen geïncorporeerde, gelabelde DNA met aclarubicine, een kankermedicijn dat intercaleert in de kleine groeven van de DNA-helix. De titratie toonde aan dat de gelabelde thymines informatie kunnen geven over plaats-specifieke binding in de context van het nucleosoom. Dit genereert nieuwe mogelijkheden voor het bestuderen van de bijdrage van nucleosomaal DNA aan nucleosoominteracties in moleculair detail.

In **Hoofdstuk 6**, bediscussiëren we het onderzoek beschreven in dit proefschrift en bieden we een breder perspectief op onze bevindingen. We bespreken potentiële richtingen voor toekomstig onderzoek aan de structurele basis van nucleosoominteracties en de rol van nucleosomaal DNA hierin in het bijzonder.

ACKNOWLEDGEMENTS

Five years and three labs have passed since I started my PhD, and now I am finally writing the last words of my thesis. There have been ups and downs, scientific inspiration and frustration, but also traveling to new and beautiful places and fun times with colleagues and friends. I would like to thank some people in particular for their contribution to this 'project' and period of my life.

First of all: Hugo, for being an inspirational daily supervisor and leader of our group. I really enjoyed being part of it. Thanks for stimulating discussions and boosting my motivation, it was crucial in the process of finishing this. Your idealism and strong belief in the importance of fundamental research is admirable and inspiring. I especially liked how we tried to figure out the architectural principles of the hypernucleosome, using foam DNA and hand-folded plastic histone models in an attempt to understand why it works the way it works. I also valued your eye for social cohesion in the group, importance of taking holidays now and then and having regular lab and team outings.

Marcellus, thank you for being my promotor for the first 2,5 years and for your valuable advice around the time of our move to Utrecht. Alexandre, thank you for taking over!

Then, of course, many thanks to the 'Hugo team' for all the discussion, support, long coffee breaks, listening to complaining and having fun in the lab and in the office over the years! Thank you Velten, tough guy and drama queen in one, for being a pioneer in how to finish a thesis in the Hugo group. I will join you now in romanticizing the freedom and quest for knowledge in science. Thank you, Ulric, for always being extremely helpful when I had any questions and for sharing the results of your Belgian baking efforts, they were delicious. Please don't forget to think about yourself now and then! Ivan, I really admire your hard work and determination to understand nucleosomal chaperones and to continue in science. Thanks for your help in the lab and critical discussion of results. Heyi, now you are up next! Thank you for sharing your feelings with us, for the philosophical discussions during coffee (about ducks, humans, robots...), and bringing some cultural perspective to the group. Vincenzo, we hand everything over to you now (no pressure). Thank you for your help on my projects, I am sure you will do great also during the second half of your PhD and after that!

A big thanks to all my Leiden colleagues, in particular Wouter, Mathilde, Antonella, Ramon, Bram, Sasha, Fredj, Soumya, Qing, Fatema and Raffaella, for scientific help and discussion, borrels at the Science club, pizza diners at VIP, D&D sessions, nerf gun fights in the old lab, making hilarious PhD videos and going on legendary lab outings. Also, thanks to my students, Yuri, Carmen and Adil, for your commitment and hard work. Although your projects on paramagnetic tagging of DNA and the HSS domain of ISWI did not end up in this thesis, they were important paths to explore. Yuri, thank you for making our lifelong stock of Pfu, I have gratefully used it in many experiments!

Thanks, also to all my new Utrecht colleagues, Alessandra, Helena, Miranda, Reinier, João, Sid, Jon, Agnes, Siri, Zuzana, Charlotte, Rhythm, Panos, Adrien, Mikael, Cecilia, Meg and Barend, for welcoming our group to your lab. I enjoyed our Friday afternoon borrels, walks in the botanical garden, ladies night, ice cream breaks and summer BBQs. Thank you all for contributing to the good time I had in Utrecht!

Importantly, I would like to thank all the supporting lab members from Leiden and Utrecht, for maintaining a functional lab, answering (stupid) questions all day and providing expert advice on experiments and weird results. Without you, no science would be possible! Thank you, Anneloes and Monika, for making the move to the new building so smooth and organized and for your positive energy! Thank you Mark and Gert, for valuable help and advice in the lab in Utrecht. Thank you, Karthick, Fons and Johan, for helping out in times of NMR crises.

The RCA and DNA labeling chapters of this thesis would not have existed without the help of Frank. Thank you so much for having me at the lab at my 'home' university in Nijmegen and sharing your seemingly endless knowledge of how to handle and manipulate DNA to make it behave the way I wanted. Thanks for being so friendly, patient, positive and helpful when I didn't know anything yet in the beginning and for solving many of the problems and setbacks that we encountered along the way.

Thank you, Johanna and Felix, for our collaboration on ISWI! I was very happy to be able to profit from your expertise and experience on the topic. Thanks also for welcoming me to your lab and to Munich, I greatly enjoyed the trip and it was crucial in boosting my confidence and motivation to continue the ISWI story at home.

Thank you, Bram, Remus, Thomas and John, for the collaboration on archaeal histones. Although it may have been challenging sometimes to understand each other's models, data and interpretation of it, we did manage to write two papers on this project in the end! I am glad I could contribute to these from a structural perspective.

To all my music friends, study friends, high school friends and childhood friends: you are all very important to me. Thanks for your support, for taking my mind off work and putting life in perspective. Olga and Elise, thank you for laughing and crying together, it means a lot. Thanks to all the people that joined me in orchestras, and to my string quartet and string trio buddies for making music together. I loved it and hope we will explore more music in the future.

Lieve familie en schoonfamilie, mam, pap, Jasper, Anne, Wouter en oom Paul, Tony en Hans, Mijke, Erwin, Luuk en Kim, dankjulliewel voor jullie steun, interesse en pogingen om te begrijpen waar ik mee bezig was. Mam en pap: zonder de stevige basis van nieuwsgierigheid, zelfvertrouwen en relativiseringsvermogen die jullie mij hebben meegegeven had ik dit niet voor elkaar gekregen.

Liefste Joris, dankjewel voor alles, je eindeloze liefde, steun en geduld. Het is eindelijk klaar nu! Ik kijk uit naar ons volgende avontuur, om binnenkort onze zoon te ontmoeten en samen onze toekomst tegemoet te gaan!

CURRICULUM VITAE

Clara van Emmerik grew up in Zwolle, the Netherlands, and completed her secondary education at the Gymnasium Ceeleum in Zwolle in 2008. She continued her studies at the Radboud University in Nijmegen completing a Bachelor's and Master's in Molecular Life Sciences. During her Master's, she was a research intern at the Institute for Genetic and Metabolic Disease (Radboudumc, Nijmegen) working on mitochondrial complex I assembly, which was followed by a six-month internship in chemical biology at the Max Planck Institute for Molecular Physiology in Dortmund, Germany. After graduating in 2014, she went to Leiden University to pursue a PhD in structural biology in the group of dr. Hugo van Ingen. For the last two years of her PhD research, she followed the team moving to the NMR spectroscopy group at Utrecht University. Since September 2019, she is a Project Leader in Product Development at MRC Holland, Amsterdam.

LIST OF PUBLICATIONS

B. Henneman, **C. van Emmerik**, H. van Ingen and R.T. Dame (2018) Structure and function of archaeal histones. *PLoS Genet*, **14**, e1007582

C.L. van Emmerik and H. van Ingen (2019) Unspinning chromatin: revealing the dynamic nucleosome landscape by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **110**, 1–19

C.L. van Emmerik, I. Gachulincova, V.R. Lobbia, M.A. Daniëls, H.A. Heus, A. Soufi, F.H.T. Nelissen and H. van Ingen (2020) Ramified rolling circle amplification for synthesis of nucleosomal DNA sequences. *Analytical Biochemistry*, **588**, 113469

B. Henneman*, T. Brouwer*, **C. van Emmerik**, R.A. van der Valk, M. Timmer, N. Kirolos, H. van Ingen, J. van Noort and R.T. Dame (2019) Mechanical and structural properties of archaeal hypernucleosomes. *Submitted*.

* These authors contributed equally.

